DOCUMENT RESUME

ED 155 760	CS 502 146
TITLE	Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for its Investigation, and Practical Application, January 1-March 31, 1978. Volumes 1 and 2.
INSTITUTION	Haskins Labs., New Haven, Conn.
REPORT NO	SR-53
PUB DATE	78
NOTE	339p.
EDRS PRICE	MF-$0.83 HC-$18.07 Plus Postage.
DESCRIPTORS	*Acoustic Phonetics; *Alphabets; Articulation (Speech); Communication (Thought Transfer); *Contrastive Linguistics; Cyrillic Alphabet; *Language Research; *Linguistics; Linguistic Theory; Metaphors; Phonology; Reading Research; Research; Research Tools; *Speech Communication; Thai; Tone Languages

ABSTRACT

This report is part of a continuing series providing information on the status and progress of studies dealing with the nature of speech, instrumentation for its investigation, and practical applications of research. The report covers the period from 1 January 1978 through 31 March 1978, and includes extended reports on the following topics: articulatory synthesis as a tool for the perceptual evaluation of articulatory gestures; the relation between processing the Roman and Cyrillic alphabets analyzed with bi-alphabetical readers; lexical decision for inflected nouns; phonetic plausibility of the segmentation of tones in Thai phonology; closure hiatus as a cue to voicing, manner, and place of consonant occlusion; metaphoric comprehension studies in reminding and resembling; skill acquisition studies in an event approach with special reference to searching for the optimum of a function of several variables; detection errors on "the" and "and"; perception of (s)+stop clusters; perceptual integration of cues for stop, fricative, and affricate manner; diachronic tone splits and voicing shifts in Thai; a range effect in the perception of voicing; perceptuomotor adaptation of speech; and interdependence of voicing and place decisions for stop consonants in initial position. Also included is a list of publications and reports, as well as Defense Documentation Center (DDC) and ERIC ordering information for earlier status reports dating from January-June 1970. (Author/DF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED155760

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SR-53, vol. 1 (1978)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 January - 31 March 1978

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

CS502146

NOTE: With Volumes 1 and 2 of SR-53 bound together, it is somewhat difficult to locate the second Table of Contents; we therefore offer the following Contents page which may be inserted into SR-53.

CONTENTS SR-53 (Volume 1)

I. Manuscripts and Extended Reports

Articulatory Synthesis - A Tool for the Perceptual Evaluation of Articulatory Gestures -- Paul Mermelstein and Philip Rubin 1

On the Relation between Processing the Roman and the Cyrillic Alphabets: A Preliminary Analysis with Bi-alphabetical Readers -- G. Lukatela, M. D. Savić, P. Ognjenović and M. T. Turvey. 13

Bi-alphabetical Lexical Decision -- G. Lukatela, M. Savić, B. Gligorijević, P. Ognjenović and M. T. Turvey 43

Lexical Decision for Inflected Nouns -- G. Lukatela, Z. Mandić, B. Gligorijević, A. Kostić, M. Savić and M. T. Turvey. 65

The Phonetic Plausibility of the Segmentation of Tones in Thai Phonology -- Arthur S. Abramson. 73

Closure Hiatus: Cue to Voicing, Manner and Place of Consonant Occlusion -- Leigh Lisker. 79

Metaphoric Comprehension: Studies in Reminding and Resembling -- Robert R. Verbrugge and Nancy S. McCarrell 87

Skill Acquisition: An Event Approach with Special Reference to Searching for the Optimum of a Function of Several Variables -- Carol A. Fowler and M. T. Turvey 127

II. Publications and Reports

III. Appendix: DDC and ERIC numbers (SR-21/22 - SR-50)

CONTENTS SR-53 (Volume 2)

I. Manuscripts and Extended Reports

Detection Errors on The and And: Evidence for Reading Units Larger than the Word -- Adam Drewnowski and Alice F. Healy 1

Some Observations on the Perception of [s]+Stop Clusters -- Peter J. Bailey and Quentin Summerfield 25

Perceptual Integration of Cues for Stop, Fricative and Affricate Manner -- Bruno H. Repp, Alvin M. Liberman, Thomas Eccardt and David Pesetsky 61

Diachronic Tone Splits and Voicing Shifts in Thai: Some Perceptual Data -- Arthur S. Abramson and Donna M. Erickson. 85

A Range Effect in the Perception of Voicing -- Susan A. Brady and Christopher J. Darwin. 97

A Note on Perceptuo-Motor Adaptation of Speech -- Quentin Summerfield, Peter J. Bailey and Donna Erickson. 105

Interdependence of Voicing and Place Decisions for Stop Consonants in Initial Position -- Bruno H. Repp 117

II. Publications and Reports

III. Appendix: DDC and ERIC numbers (SR-21/22 - SR-50)

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994

Assistant Chief Medical Director for Research and Development, Research Center for Prosthetics, Veterans Administration
Contract V101(134)P-342

National Institutes of Child Health and Human Development
Contract N01-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-5596

National Science Foundation
Grant BNS76-82023

National Science Foundation
Grant MCS76-81034

National Institute of Neurological and Communicative Disorders and Stroke
Grant NS13870
Grant NS13617

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper, Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey, Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Thomas Baer
Fredericka Bell-Berti
Gloria J. Borden*
Guy Carden*
Robert Crowder*
Steven B. Davis
Michael Dorman*
Donna Erickson
William Ewan*
Carol A. Fowler*
Frances J. Freeman*
Jane H. Gaitenby
Thomas J. Gay*
Katherine S. Harris*
Alice Healy*
David Isenberg
Leonard Katz*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann
Ignatius G. Mattingly*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Donald P. Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Robert Verbrugge*
Hirohide Yoshioka[1]

Technical and Support Staff

Eric L. Andreasson
Elizabeth P. Clark
Donald Hailey
Terry Halwes
Elly Knight*
Sabina D. Koroluk
Agnes McKeon
Nancy R. O'Brien
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
Bobbe Wire
David Zeichner

Students*

Linda D'Antonio
William Balch
Steve Braddon
David Dechovitz
Laurel Dent
Susan Lea Donald
F. William Fischer
Hollis Fitch
Carole E. Gelfer
Janette Henderson
Nieba Jones
Morey J. Kitzman
Andrea G. Levitt
Roland Mandler
Leonard Mark
Nancy McGarr
Georgia Nigro
Mary Jo Osberger
Abigail Reilly
Arnold Shapiro
Emily Tobey
Betty Tuller
Harold Tzeutschler
Michele Werfelman

---

*Part-time
[1] Visiting from University of Tokyo, Japan.

CONTENTS

I. Manuscripts and Extended Reports

Articulatory Synthesis - A Tool for the Perceptual Evaluation of Articulatory Gestures -- Paul Mermelstein and Philip Rubin 1

On the Relation between Processing the Roman and the Cyrillic Alphabets: A Preliminary Analysis with Bi-alphabetical Readers -- G. Lukatela, M. D. Savić, P. Ognjenović and M. T. Turvey 13

Bi-alphabetical Lexical Decision -- G. Lukatela, M. Savić, B. Gligorijević, P. Ognjenović and M. T. Turvey 43

Lexical Decision for Inflected Nouns -- G. Lukatela, Z. Mandić, B. Gligorijević, A. Kostić, M. Savić and M. T. Turvey 65

The Phonetic Plausibility of the Segmentation of Tones in Thai Phonology -- Arthur S. Abramson 73

Closure Hiatus: Cue to Voicing, Manner and Place of Consonant Occlusion -- Leigh Lisker 79

Metaphoric Comprehension: Studies in Reminding and Resembling -- Robert R. Verbrugge and Nancy S. McCarrell 87

Skill Acquisition: An Event Approach with Special Reference to Searching for the Optimum of a Function of Several Variables -- Carol A. Fowler and M. T. Turvey 127

II. Publications and Reports

III. Appendix: DDC and ERIC numbers (SR-21/22 - SR-50)

I. MANUSCRIPTS AND EXTENDED REPORTS

Articulatory Synthesis - A Tool for the Perceptual Evaluation of Articulatory Gestures[*]

Paul Mermelstein[†] and Philip Rubin

ABSTRACT

Over the last twenty-five years, acoustic speech synthesis from spectrally specified parameters has served as a unique tool in assessing the perceptual importance of acoustic features present in the speech signal. Articulatory features have met with less attention, perhaps because they cannot be directly observed in a spectrogram. On the static level, articulatory synthesis made it possible to study the acoustic consequences of varying the position of independent articulators. However, such static representations are not wholly adequate from the perceptual point of view. For example, the identification of isolated vowels is a perceptually more difficult task than the identification of vowels in a syllable environment. A significant body of evidence leads us to believe that the listener uses knowledge about constraints on the production mechanism to interpret speech stimuli. Articulatory synthesis appears to be an ideal tool for exploring those dynamic aspects of the articulatory process that convey information that a listener may employ in phonetic processing. The development of such a synthesizer into a useful research tool is outlined.

INTRODUCTION

The purpose of this paper is to identify an area of investigation in which the use of an articulatory synthesizer can be expected to contribute to the understanding of speech perception by supporting experimental methodologies that have rarely been employed in the past. Paralleling the research that has been continuing at Haskins Laboratories and other research institutions for many years, we intend to use the synthesizer as a tool to examine the nature of perceptually significant articulatory information. Articulatory

---

[*] Portions of this paper were presented at the Symposium on Articulatory Modelling, Grenoble, France, 11-12 July, 1977.

[†] Also at Bell-Northern Research and INRS-Telecommunications, University of Quebec, Montreal, Canada.

Acknowledgment: The authors wish to express their appreciation to Patrick Nye and Thomas Baer for their advice and criticisms on earlier drafts of this manuscript. Preparation of this paper was supported by NSF Grant BNS-76-82023, and BRSG Grant RR-05596, to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 1 (1978)]

synthesis allows the experimenter to precisely control speech gestures by specifying the positions of selected articulatory variables, incorporating these variables into programs for generating the articulatory trajectory, and testing whether listeners in fact regard these variables as important for the perception of particular speech sounds. With such a synthesizer, the experimenter can form an hypothesis about which gestures are significant, incorporate variations of these gestures into programs for generating the articulatory trajectory, and test whether listeners regard these variations as important for the perception of particular speech sounds.

USE OF ARTICULATORY SYNTHESIS

We have already pointed out that the use of articulatory synthesis to probe human speech perception represents a method of research that parallels the use of spectrally-based synthesis to determine the perceptually important acoustic features. The organization of this research method is illustrated in Figure 1, which shows three ways to experiment on speech.[1] These are: (1) the use of a real speaker, (2) an articulatory model and (3) a terminal analog model. When real speech is used as input, a digital playback device can serve to analyze the signal into its spectral components, possibly manipulate the spectrographic representation and resynthesize the signal. An articulatory model generates synthetic speech that may be compared to real speech by listening, that is, subjected to perceptual evaluation. More importantly, however, the control signals of the muscles of the articulators that are observed through electromyographic (EMG) measurements can be compared with the signals that drive the articulatory model. Unfortunately, the signals cannot be compared quantitatively at the moment, except in terms of timing information.

At the vocal-tract shape level, a comparison is possible between sagittal x-ray views and the schematized displays of the articulatory model. In addition, the spectrograms resulting from articulatory synthesis can be compared with the spectrograms obtained through speech synthesis using a terminal analog synthesizer. This permits us to verify whether the perceived differences in the synthetic speech signals are due to a failure to specify the proper acoustic information adequately, or whether they are more likely the consequence of incorrect articulatory specifications.

THE MODEL

The details of the particular articulatory synthesis model that we have implemented as a first step have been described previously by Mermelstein (1973). The positions of six key articulators are controlled. These articulators can be divided into two groups: (a) primary - those that move independently of the other articulators; and (b) secondary - those whose positions are functions of the position of other articulators.

---

[1] Figure 1 is reproduced from "Speech Synthesis - A Tool for the Study of Speech Production" by F. S. Cooper, P. Mermelstein and P. W. Nye, in Dynamic Aspects of Speech Production - Current Results, Emerging Problems and New Instrumentation, edited by M. Sawashima and F. S. Cooper. (Tokyo: University of Tokyo Press, 1977).

The jaw, velum and hyoid constitute the first group: the tongue body, tongue tip and lips constitute the second group. The articulators of the second group all move relative to the jaw. When articulatory movements are modeled in this manner, individual speech gestures can be separated into the component movements of each of the several articulators that are involved. For example, the lip opening gesture in articulating /ba/ has two main components; the release of lip closure and the opening of the jaw for the vowel articulation. Movements of the jaw and velum have one degree of freedom, all other articulators move with two degrees of freedom.

EXERCISING CONTROL OVER THE ARTICULATORY SYNTHESIZER

The articulatory process is simulated on a digital computer. Digital simulation can provide a flexibility and convenience that is unattainable through the use of physical models. To control the model meaningfully, it must be possible to easily observe the results of changes in the input instructions. To this end, a graphical display, as shown in Figure 2a, is provided that allows the user to select an individual articulator and move it to a specified position. The vocal-tract outline is immediately recalculated and the modified display is made available for inspection. Once excitation parameters are specified, the model calculates the transfer function from the specified vocal-tract shape and displays the appropriate spectrum, whether voiced or fricative. Finally, the model generates an acoustic output by computing a digital representation of the soundwave from the transfer function. To examine stationary vocal-tract configurations, a standard descending fundamental frequency trajectory is synthesized for a duration of 200 msec. This stationary mode is used primarily to evaluate changes in the vowel color resulting from perturbations in the specification of particular vocal-tract shapes.

To generate articulatory movements, an input consisting of a sequence of articulatory states is provided by the user. This set of specifications takes the form of two tables of values. The first table, referred to as the "script" table, consists of descriptions of the positions of the articulators within the vocal-tract at particular points in time (see Figure 5). Therefore, each row of a script table describes a particular shape of the vocal-tract. A second table, called the "control" table, controls the timing, fundamental frequency and amplitude parameters, and specifies, if necessary, the point in the vocal tract where the fricative noise source is to be introduced (see Figure 7). The use of these tables is similar to the procedure known as key-frame animation. Key "snapshots" of the vocal tract are provided in a particular order by the script table. The flow of movement is determined by interpolating, or moving, between these critical articulatory states, as specified by the timing parameters in the control table. The result, then, is a simulation of movement, or animation, of the vocal tract through a path of key configurations. At the moment, each articulatory parameter value is linearly interpolated between the values specified in the script table. Future modifications will allow the experimenter to specify exponential transitions with variable time constants.

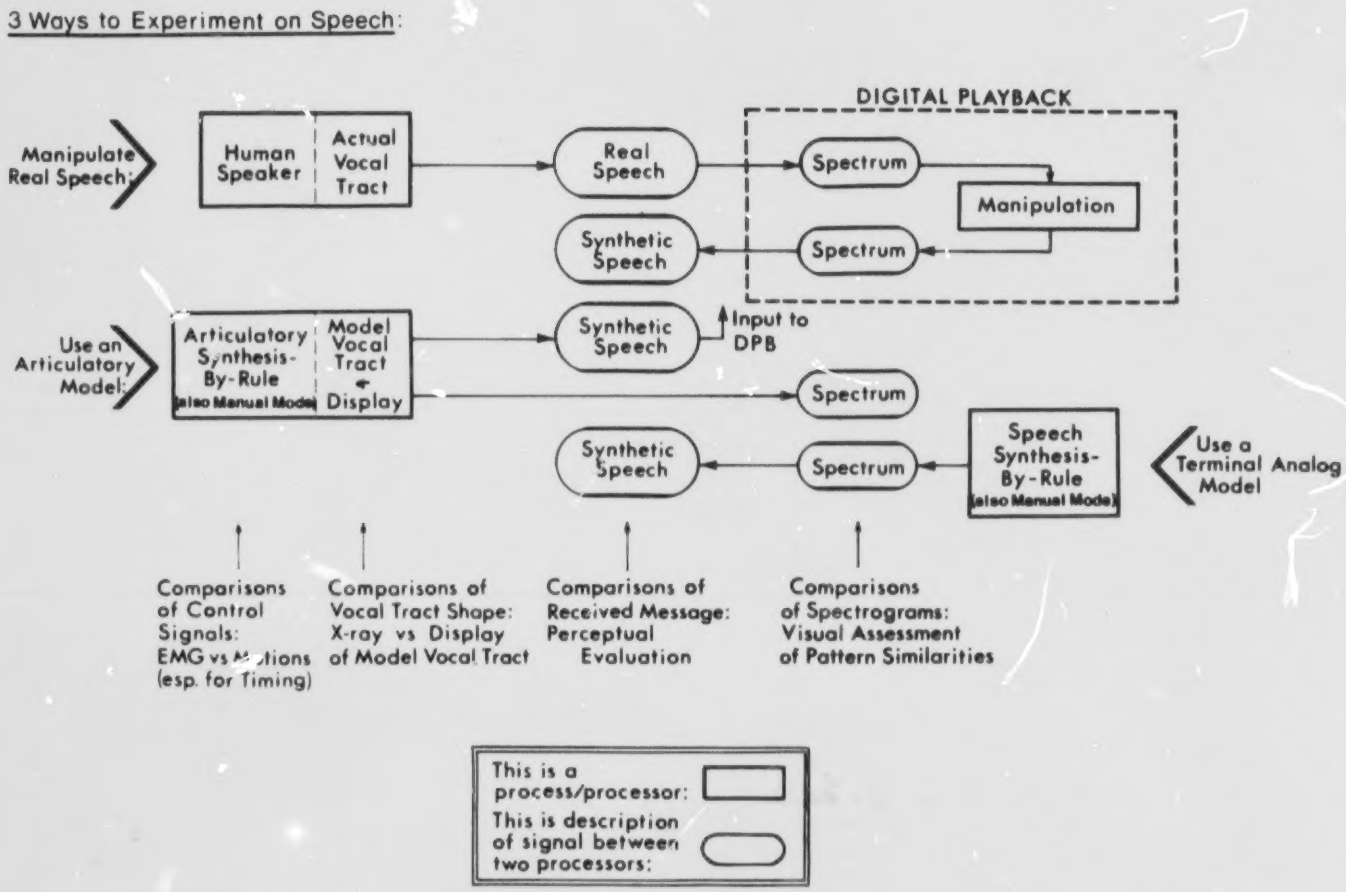

3 Ways to Experiment on Speech:
DIGITAL PLAYBACK
Manipulate Real Speech:
Human Speaker
Actual Vocal Tract
Real Speech
Spectrum
Manipulation
Synthetic Speech
Spectrum
Use an Articulatory Model:
Articulatory Synthesis-By-Rule (also Manual Mode)
Model Vocal Tract ← Display
Synthetic Speech
Input to DPB
Spectrum
Synthetic Speech
Spectrum
Speech Synthesis-By-Rule (also Manual Mode)
Use a Terminal Analog Model
Comparisons of Control Signals: EMG vs Motions (esp. for Timing)
Comparisons of Vocal Tract Shape: X-ray vs Display of Model Vocal Tract
Comparisons of Received Message: Perceptual Evaluation
Comparisons of Spectrograms: Visual Assessment of Pattern Similarities
This is a process/processor:
This is description of signal between two processors:

Figure 1: Three ways to experiment on speech.

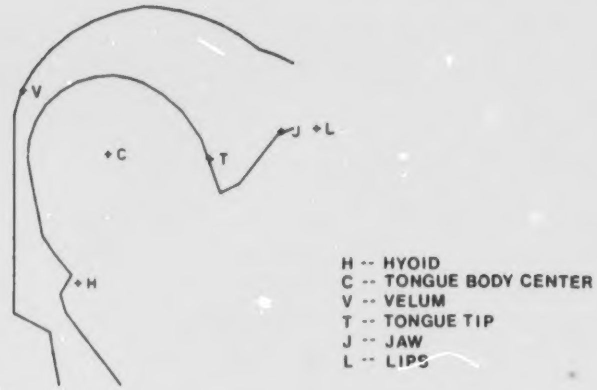

V
C
T
J
L
H
H -- HYOID
C -- TONGUE BODY CENTER
V -- VELUM
T -- TONGUE TIP
J -- JAW
L -- LIPS

Figure 2a: Articulatory display for vowel /a/.

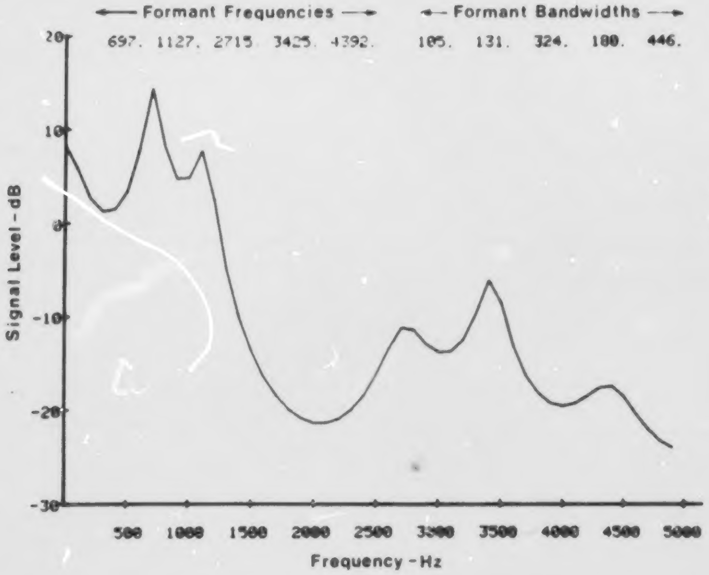


Figure 2b: Vocal-tract transfer function for vowel /a/.

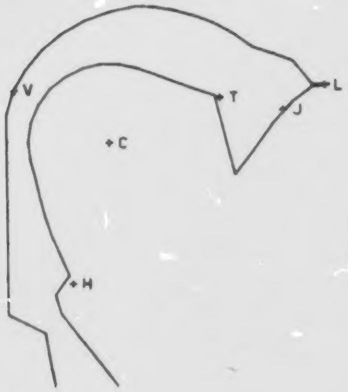


Figure 3a: Articulatory display for stop /b/.

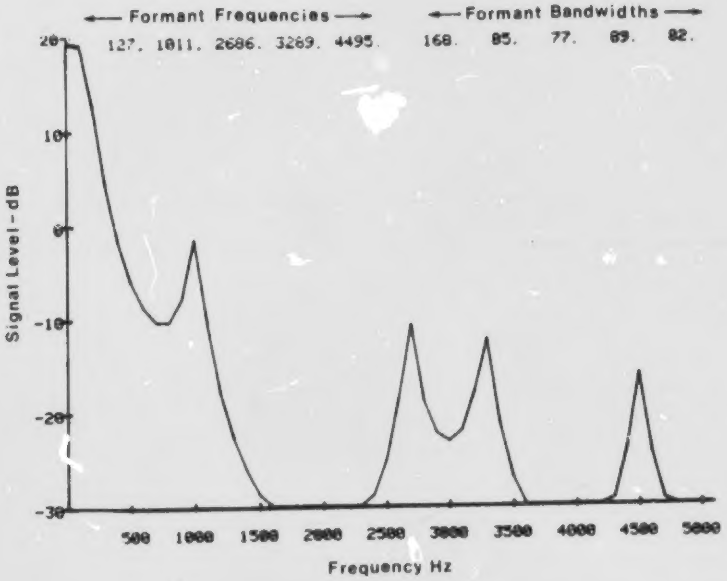


Figure 3b: Vocal-tract transfer function for stop /b/.

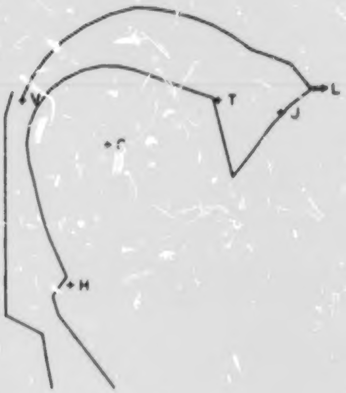


Figure 4a: Articulatory display for nasal /m/.

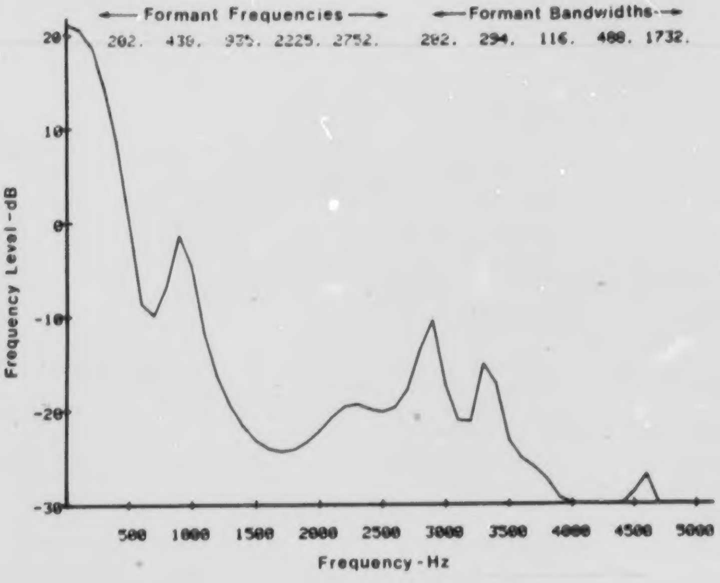


Figure 4b: Vocal-tract transfer function for nasal /m/.

```
TABLE OF ARTICULATORY PARAMETERS WITH DEFAULT VALUES:

     H-X     H-Y     SC      THC     ST      THT     THJ     L-P     L-H     NAS
     800     830     856     -.21    350     0.      -.28    102     11      0.0
---------------------------------------------------------------------------------
1:   800     830     845.7   -.278   303.1   0.38    -.299   119.9    70.9   0.0
2:   800     830     845.7   -.278   303.1   0.38    -.299   119.9    70.9   0.0
3:   800     830     845.7   -.278   350.0   0.00    -.346   102.9     9.1   0.0
4:   800     830     845.7   -.278   350.0   0.00    -.346   102.9     9.1   0.0
5:   800     830     845.7   -.278   350.0   0.00    -.346   102.9     9.1   0.0
6:   800     830     845.7   -.278   350.0   0.00    -.346   102.9     9.1   0.0
```

H-X -- hyoid position, X coordinate
H-Y -- hyoid position, Y coordinate
SC -- distance from origin to tongue body center
THC -- angle (in radians) between jaw and tongue body
ST -- tongue tip extension - position relative to tongue body
THT -- tongue tip angle (in radians)
THJ -- angular position of jaw relative to horizontal
L-P -- lip protrusion
L-H -- lip height
NAS -- velum height; velar port opening

Figure 5: Articulation script for /ba/.

TABLE OF ARTICULATORY PARAMETERS WITH DEFAULT VALUES:

	H-X	H-Y	SC	THC	ST	THT	THJ	L-P	L-H	NAS
	800	830	856	-.21	350	0.	-.28	102	11	0.0
1:	800	830	845.7	-.278	303.1	0.38	-.299	119.9	70.9	0.045
2:	800	830	845.7	-.278	303.1	0.38	-.299	119.9	70.9	0.045
3:	800	830	845.7	-.278	350.0	0.00	-.346	102.9	9.1	0.045
4:	800	830	845.7	-.278	350.0	0.00	-.346	102.9	9.1	0.000
5:	800	830	845.7	-.278	350.0	0.00	-.346	102.9	9.1	0.000
6:	800	830	845.7	-.278	350.0	0.00	-.346	102.9	9.1	0.000

Figure 6: Articulation script for /ma/.

TABLE OF ARTICULATORY SYNTHESIS CONTROL PARAMETERS WITH DEFAULT VALUES:

	TIME	AMP	AMPFR	NFRICP	FREQ
	0	0	0	0	100
1:	0.0	20.0	0.0	0.	120.
2:	150.0	20.0	0.0	0.	120.
3:	200.0	20.0	0.0	0.	120.
4:	240.0	20.0	0.0	0.	120.
5:	350.0	20.0	0.0	0.	90.
6:	375.0	0.0	0.0	0.	85.

TIME -- starting time of a table row (msec)
AMP -- input voicing amplitude (arbitrary scale)
AMPFR -- input frication amplitude (arbitrary scale)
NFRICP -- point in the vocal tract where noise source is inserted (from larynx to lips)
FREQ -- fundamental frequency (Hz)

Figure 7: Timing and excitation control for /ba/ and /ma/.

Since it is difficult to visualize articulatory shapes when they are specified in numeric form, the program provides a graphical display of the shape specified at any time in the sequence of articulatory gestures. Modifications to these vocal-tract shapes can then be carried out graphically in the stationary mode and the numerical results can be automatically substituted back into the script table.

Movements of the vocal tract are not simulated continuously. The positions of the articulators are determined at the onset of every pitch period and the corresponding transfer function is computed. The resulting speech signal is obtained by concatenating the truncated responses to individual pitch pulses of varying durations.

Although the acoustic signal is not computed in real time, it is generally produced within no more than fifty times real time. Prompt output is desirable since it allows the user to quickly assess the perceptual consequences of the synthesis process. It is only under such conditions of rapid feedback that the user can maintain a conceptual link between the hypothesis being tested and the results of the test.

The following figures illustrate the input-output relationships of the model at the transfer-function level. Figure 2a shows the spatial positions of the key articulators involved in the formation of a vocal tract outline appropriate for the production of the vowel /a/. Figure 2b shows the corresponding transfer function. The pole frequencies and bandwidths listed at the top of Figure 2b are determined by solving for the roots of the denominator of the transfer-function polynomial. To generate /ba/, the vowel articulation is preceded by a vocal-tract outline with closed lips as shown in Figure 3a. The corresponding transfer function is shown in Figure 3b. Because a small opening at the lips is being used to simulate radiation through the cheeks, the higher formant bandwidths tend to be too small. Figure 4a is an articulatory configuration appropriate for the consonant /m/, requiring articulatory specifications for velar opening and labial closure. The corresponding transfer function is shown in Figure 4b. Figure 5 illustrates a typical script table, this one appropriate for /ba/. The changes in the tongue-tip coordinates are not important. Rather, it is the changes in jaw and lip parameters that are noteworthy. The specification for /ma/, as seen in Figure 6, is identical to that of /ba/, except for the specification of the velar parameter. Figure 7 illustrates the corresponding control table where timing and excitation parameters are specified.

APPLICATIONS

The research issues that we hope to explore with the aid of the articulatory model revolve around the identification of the articulatory components of a vocal gesture that are perceptually important. In the stationary mode, that is, when listening to or comparing sustained speech sounds, it is difficult to specify whether one is perceiving in an acoustic or in an articulatory framework. For time varying speech sounds, such as vowels in CVC contexts, the situation may be very different. We have previously found that when displacing one of the formant frequencies of a vowel, the just noticeable difference (JND) is significantly smaller if the vowel being modified is the central vowel of a CVC syllable, than when the vowel is stationary (Mermelstein, 1977). The JND is increased even in cases where the particular formant frequency being perturbed does not normally vary with time. The need to decode the consonantal information appears to prevent the listeners from discriminating differences between the vowels as well as they can when the consonants are absent. However, the JND for a stationary formant is smaller than the JND for a formant that is varying in time. It appears that the increase in the JND for formant frequency has partially an auditory and partially a phonetic basis.

It has been suggested that vowels heard in context are easier to identify than vowels heard in isolation, because coarticulation between a consonant and a following vowel causes the consonant to carry some information about the vowel as well (Strange et al., 1976). Hence, in a syllabic context, information pertaining to the vowel is available not only from the nuclear region of the syllable, but also from the consonantal environment. To put this hypothesis into a form testable in articulatory terms, we may ask: Is the JND in position for an articulator (measured at a moment when it is most representative of the vowel uttered alone) reduced when the articulator participates in the consonantal movement as well? Presumably, under such dynamic conditions, more information about the articulator's intended position is available to the listener. Hence, to be specific, we may ask whether the JND in lip opening depends on whether the vowel is in a labial or a velar context. A reduced value for the JND in a labial context would suggest that this context does provide some assistance to the listener in assessing the identity of an adjoining vowel

Another area that we plan to explore with the aid of the articulatory model is the perceptual sensitivity of listeners to variations in the timing of overlapped articulatory gestures. Certain articulatory events appear to be precisely time-locked, and we suspect that any disturbance of that natural precision could be perceptually disruptive. However, based on the examinaton of repeated productions, other events appear to be less governed by uniform timing constraints. Nevertheless, the degree of perceptual awareness of these differing requirements has not yet been demonstrated.

The answers to such questions about the perceptual significance of various components of articulatory performance promise to shed new light on the speech perception process. Moreover, the development of an articulatory synthesizer as a research tool has made it possible to study the assumed link between speech perception and production (for example, Liberman et al., 1967) in a more feasible and revealing way than has been possible in the past.

REFERENCES

Liberman, A. M., F. S. Cooper, D. P. Shankweiler and M. Studdert-Kennedy. (1967) Perception of the speech code. Psychological Review 74, 431-461.

Mermelstein, P. (1973) Articulatory model for the study of speech production. Journal of the Acoustical Society of America 53, 1070-1082.

Mermelstein, P. (in press) Difference limens for formant frequencies of steady-state and consonant-bounded vowels. Journal of the Acoustical Society of America.

Strange, W., R. R. Verbrugge, D. P. Shankweiler and T. R. Edman. (1976) Consonant environment specifies vowel identity. Journal of the Acoustical Society of America 60, 213-224.

On the Relation between Processing the Roman and the Cyrillic Alphabets: A Preliminary Analysis with Bi-alphabetical Readers[*]

G. Lukatela[†], M. D. Savić[†], P. Ognjenović[†] and M. T. Turvey[††]

ABSTRACT

Serbo-Croatian is read, to a greater or lesser degree depending on locale, in two alphabets, the Roman and the Cyrillic. While most letters are solely members of one or the other alphabet, some letters are shared and of these, some are ambiguous in that they are read differently in the two alphabets. The order in which the alphabets are acquired depends on geography: in the eastern part of the country the order is Cyrillic then Roman; in the western part of the country the order is Roman then Cyrillic. A series of six experiments is reported examining the relation, in processing terms, between the two alphabets. Evidence is presented for a processing asymmetry. Processing the letters of the first-acquired alphabet is more similar to processing the letters of the second-acquired alphabet than vice versa. Additionally, it is shown that searching for a letter in the other alphabet is faster than searching for a letter in the same alphabet, suggesting that alphabet categorization may precede letter identification. The results are discussed in terms of the general problem of operating with two separately used symbol systems.

INTRODUCTION

The modern Serbo-Croatian orthography was constructed at the beginning of the 19th century by Vuk Karadić on the basis of a simple rule: "Write as you speak." He selected the speech spoken in mid-Yugoslavia as the ideal, and to each phonemic segment of the speech he assigned a letter character. Karadić took the majority of the letters from the alphabet existing at the time, but since the number of letters available was less than the number of phonemes needed, he borrowed and/or modified several letters from other alphabets. Consequently, in the modern alphabet, each letter stands for a phoneme and the phonemic interpretation of each individual letter is largely invariant and unaffected by preceding and following letters and letter clusters. All

---

[*] To appear in Language and Speech, vol. 21, 1978.

[†] University of Belgrade.

[††] Also University of Connecticut.

Acknowledgment: This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade, and in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 1 (1978)]

letters are pronounced; there are no letters that are made silent by context.

In actuality, there are two alphabets with the above properties--a Roman and a Cyrillic--and in many areas of Yugoslavia both are used by the local population. This situation is due, in part, to the educational system that teaches both alphabets in the first and second grade and, in part, to the fact that reading materials come in both alphabets. In Eastern Yugoslavia the children are taught to read and write Cyrillic during their first school year and Roman during their second; in Western Yugoslavia the children learn first Roman and then Cyrillic. Consequently, the normal third grade child in most of Yugoslavia can handle both alphabets.

The Cyrillic and Roman alphabets in Serbo-Croatian do not represent two completely independent sets of letters. Serbo-Croatian letters can be divided into four different groups, which are illustrated in Figure 1. Some letters are the same in shape and pronunciation in both alphabets (see Table 1 for the pronunciations). We will refer to these letters as "common letters." The word for aunt, for example, is written TETKA in Roman and in Cyrillic. However, there are also several letters of the same shape that represent, in the two alphabets, different utterances. We will call them "ambiguous letters." The word deer, for example, is spelled CPHA in Cyrillic. However, if CPHA were read as Roman, the pronunciation would be different and the "word" itself would be meaningless. Similarly, one can combine ambiguous and common letters to write words which have one pronunciation and meaning if read as Cyrillic, and a different pronunciation and a different meaning if read as Roman. Finally, the remaining letters are specific either to the Roman or Cyrillic alphabets; we will refer to these as "the uniquely Roman" or "the uniquely Cyrillic" letters, respectively.

It is evident that the relation between the two alphabets is not the same as the relation between the upper- and lower-case alphabets of, say, English. It is also evident from the preceding that Serbo-Croatian provides a special situation for the study of word perception in particular, and reading in general. Our initial interest, however, is with an issue that is more modest than, and perhaps preliminary to, the larger issues of word perception and reading, namely: What is the relation, in processing terms, between the two alphabets? The present paper reports six experiments that bear on this problem.

Let us preface these experiments with some general comments about the learning of the two alphabets. Fundamentally, alphabetic characters are visual specifications of articulatory events; each character specifies a unique speech sound. Nevertheless, differentiating the written characters, one from another, must logically precede decoding them to speech (Gibson, 1965). At the outset, then, learning an alphabet is a matter of distinguishing among a set of line-complexes that are alike on some dimensions of description and different on other dimensions of description. Sensitivity to the dimensions of difference is the initial goal. This is not a trivial requirement, since the dimensions of difference (which, for simplicity, can be called features) are probably relational so as to remain invariant under the variety of metrical and affine transformations to which writing necessarily subjects them.

If representations (such as templates) of the individual characters are induced in the course of acquiring the alphabet, then it is reasonable to suppose that the dimensions of difference constitute the descriptors from which the representations are composed. In short, differentiation of alphabet characters must, in all probability, precede representation of alphabet characters (see Gibson, 1969). Presumably, the induction of the mapping from the characters to speech is possible once a reasonable level of distinction among the characters has been achieved and representation made feasible.

As remarked, Yugoslavians indigenous to Eastern Yugoslavia learn the Cyrillic alphabet first. On the foregoing, this means that they have learned to detect the dimensions of difference relevant to the set of Cyrillic characters; they have acquired, presumably, representations for the individual Cyrillic characters; they have isolated the subset of articulatory events to which the characters correspond and they have established the correspondences. What, then, does learning of the second alphabet, the Roman, require? First, we may ask: Are the dimensions of difference for the set of Roman characters the same as for the set of Cyrillic characters? Casual inspection of Figure 1 suggests that there are probably features related to distinguishing Cyrillic (in particular, the uniquely Cyrillic) characters that are irrelevant to distinguishing Roman characters, but that a subset of the Cyrillic-relevant features will probably do for the task of distinguishing Roman characters. Second, learning the Roman alphabet would not require the isolation of the relevant subset of articulatory events. Third, it is evident that the full complement of correspondencies between Roman characters and speech does not have to be learned, since seven of the Roman characters are shared with the Cyrillic alphabet. The common letters yield perfect positive transfer. In contrast, the ambiguous letters--those that are the same in shape but correspond to different speech sounds in the two alphabets--yield very high negative transfer and would require exceptional attention in the acquisition of the Roman alphabet. In this respect it is noteworthy that the elementary schoolchild, having previously learned Cyrillic, is often admonished: "Remember, you are now reading Roman."

Simplistically, there are two characterizations of the way in which the learning of the two alphabets might proceed. One characterization is that, figuratively speaking, two separate devices are constructed: the first one to accept the Cyrillic alphabet and the second to accept the Roman alphabet. Let C and R, respectively, designate the two devices. In the other characterization, a device is constructed to accept the Cyrillic alphabet and then modifications to this device are discovered so that the Cyrillic-alphabet acceptance device, suitably modified, accepts the Roman alphabet. If C designates the Cyrillic-alphabet device, then m(C) designates the modified device for accepting Roman. In view of the preceding discussion on the successive learning of two alphabets, the second characterization seems the more likely of the two. Significantly, the two characterizations are nontrivially distinct. The second implies that while processing Roman characters necessarily entails the device for processing Cyrillic characters, the reverse is not true. That is, m(C) entails C, but C does not entail m(C). In contrast, the first characterization does not imply the entailment of one alphabet device by the other, asymmetric or otherwise.

EXPERIMENT I

The first experiment sought to provide some rudimentary data of relevance to the question of how the two alphabets relate. The experiment was simple in conception and implementation: it asked native Eastern Yugoslavians to look at Roman and Cyrillic letters presented one at a time in random order and to press a key as quickly as possible in answer to the question "Is this letter Cyrillic?" or to the question "Is this letter Roman?"

Method

Subjects. The participants in the experiment were 38 students from the Psychology Department at the University of Belgrade. The students had all received their elementary education in Eastern Yugoslavia. They were experienced in reaction time experiments.

Materials. The letters were Letraset, black uppercase letters (Helvetica Light, twelve point). They were presented on slides, one letter per slide located at the center. Of the uniquely Cyrillic letters, all were used with the exception of Y and N. Of the uniquely Roman letters, those excluded were U and I, the Roman equivalents of Y and N, and those letters of the Roman alphabet that are truly combinations of letters, namely, DJ, NJ and DZ. Also excluded were three common letters: A, E and O (see Figure 1). The resulting 39 letters were divided into the following classes: ambiguous letters, common letters, uniquely Cyrillic letters and uniquely Roman letters.

Design. Each subject was assigned by order of appearance to one of two groups, with nineteen subjects per group. Both groups saw the full complement of Roman and Cyrillic letters. One group was instructed to respond "yes" or "no" to the question "Is this letter Cyrillic?"; the other group was instructed to respond similarly to the question "Is this letter Roman?"

Each subject viewed and responded to a total of 144 slides, with each letter appearing at least three times. Within a block of 36 slides the four groups of letters were quasi-randomly presented. The constraint was that no more than four letters from the same group could occur in succession. Within a block of 36 slides "Yes" and "No" responses occurred equally often.

Procedure. The letters were presented each for 200 msec in one field of a Scientific Prototype three-channel tachistoscope with another field providing a point of fixation prior to exposure. The luminances of the two fields were matched at 10.3 cd/m^2.

The onset of a letter display triggered an electronic counter that was stopped when the subject pressed one of two keys on the response panel in front of him. To minimize possible hand asymmetries, both hands were used: both thumbs were placed on the key close to the subject and both forefingers were placed on the key that was collinear with the first, but two inches further away. The subject depressed the closer key for "no" and the farther key for "yes." The duration of a display was terminated by the key press.

All subjects received ten minutes of practice preliminary to the experiment proper. After every block of 36 trials there was a brief rest period.

Results

Only correct responses were analyzed. The error rate in the two nonambiguous-letter classes was less than three percent; in the ambiguous letter and common letter classes the error rates were closer to eight percent.

The mean reaction times for each letter within a class were averaged across the subjects and then the class average was determined. The results are given in Figures 2 and 3.

Inspection of the aforementioned figures reveals that the subjects behaved quite differently under the two question regimes. First, we may note that it took considerably longer to verify that the common letters were Roman in the "Is this letter Roman?" condition than to verify that the common letters were Cyrillic in the "Is this letter Cyrillic?" condition. To dramatize this contrast we plot (in Figure 4) the probability of occurrence in Serbo-Croatian literature of each of the common letters (Tomić, 1975) against their verification latencies in the two question regimes. There is a suggestion that verification latency is an inverse function of probability of occurrence, but that the superiority of verification in the Cyrillic mode over that in the Roman mode is indifferent to a letter's probability. By an independent t-test, the latency difference between the two modes for the class of common letters was shown to be significant ($t = 3.3$, $df = 6$, $p < .02$).

Second, we may note that while there is no difference between the two question regimes when the class of letters is nonambiguous and the response is "yes," there is a substantial difference between the two for that class of letters when the response is "no." In short, the subjects accepted an unambiguous Roman letter as "Roman" and an unambiguous Cyrillic letter as "Cyrillic" with equal facility, but found it inordinately more difficult to reject a Cyrillic letter as Roman than to reject a Roman letter as Cyrillic. From the set of 14 nonambiguous Roman letters and 17 nonambiguous Cyrillic letters, 10 pairs can be identified that are phonemically equivalent. An independent t-test on the latencies for rejecting these 10 Roman letters as Cyrillic and rejecting the corresponding 10 Cyrillic letters as Roman, proved significant ($t = 3.35$, $df = 36$, $p < .01$).

Third, and last, it can be observed from Figures 2 and 3 that verifying that the ambiguous letters were members of the Cyrillic alphabet and verifying that they were members of the Roman alphabet took virtually the same amount of time. In both cases, however, these verifications were slower than those for the uniquely Cyrillic or uniquely Roman letters ($t = 5.1$, $df = 18$, $p < .01$ and $t = 2.7$, $df = 18$, $p < .05$ respectively).

Discussion

The alphabet classification task of this experiment is not a natural one. The reader of Serbo-Croatian uses his knowledge of the alphabets to go from script to meaning, but he does not ask himself--at least not explicitly--whether this or that letter is Roman or Cyrillic. Nevertheless, the task ought to reveal something of the structure of the reader's alphabet system--much as the lexical decision task ("is this string of letters a word or not?") and its variants have cast some light on the structure of the lexicon (for example, Forster and Bednall, 1976).

Figure 5 depicts one way of conveying the flavor of our introductory remarks on the learning of the two alphabets. The solid lines identify the initially established device, C, for accepting the Cyrillic alphabet, and the dotted lines identify the modifications to C that produce m(C), the device for accepting Roman. The intersection of the two alphabet spaces is the set of representations of common letters. Collectively, the two devices might operate in the following fashion. On presentation, a letter's feature description is determined and then mapped onto the two alphabet spaces in serial or in parallel. Where a match is made between the letter's figural description and that registered in one or the other visual space, an alphabet classification is defined. The accessing of the phonemic space (and other linguistic spaces) is made only subsequent to such a match. We remain uncommitted on the level at which context influences processing: if the mapping from feature description to alphabet space is serial, then context (for example, "is this character Roman?") may direct this mapping; on the other hand, if the mapping from feature description to alphabet space is parallel, then context may direct the subsequent mapping of alphabet representations onto the phonemic space. However, we need not necessarily believe that context effects are the exclusive prerogative of any one level of processing.

A major conclusion of the present experiment is that the participants viewed the common letters as essentially members of the Cyrillic alphabet and, perhaps, only indirectly as members of the Roman alphabet. This bias toward Cyrillic is not especially surprising when one considers that the subjects received their elementary education in Eastern Yugoslavia and thus learned Cyrillic as their first alphabet. The bias is surprising, however, when one recognizes that the subjects were senior university students who spend most of their (academic) reading and writing lives with the Roman alphabet.

The shorter latency for accepting common letters as Cyrillic suggests that--for these subjects--to perceive a common letter is to operate in the Cyrillic alphabet space and to conclude that a common letter is indeed Roman requires further processing of a more contrived nature. In short, to read common Roman characters it is only necessary that the representations of the common letters be accessible; it is not necessary that they be identified explicitly, within the system, as Roman.

What of the ambiguous letters? We can conjecture that they inhabit both the Roman and the Cyrillic alphabet spaces. Thus, given an ambiguous letter, a match can be found in both alphabet spaces, and for a subsequent decision process there is reason for hesitancy. In both question regimes of the present experiment, verifying that an ambiguous letter was a member of the designated alphabet took significantly longer than verifying the alphabetic membership of a letter that belonged to only one alphabet. By itself, the necessity to keep the ambiguous letters from mutually interfering suggests that the Yugoslav reader indulges two alphabet spaces and, as a consequence, he or she can be said to read in one alphabet mode or the other.

Let us now consider what is, perhaps, the most telling observation of the present experiment: rejecting Cyrillic letters in the Roman mode takes longer than rejecting Roman letters in the Cyrillic mode. To begin with, this observation rules out a simple interpretation of the relation between processing Cyrillic characters and processing Roman characters. In view of the aforementioned Cyrillic bias on common letters, it would be argued that the Cyrillic space is the larger of the two in that it contained more elements (uniquely Cyrillic, ambiguous and common versus uniquely Roman and ambiguous). Now we could imagine that when asked "Is this letter Roman?" or "Is this letter Cyrillic?", the participant engages in a search of the appropriate space looking for a match. In the case where the target is not in the specified alphabet, we may assume that the search is exhaustive (see Forster and Bednall, 1976). Therefore, if the Roman is the smaller alphabet space, then the time to reject a nonentry (a uniquely Cyrillic letter) in the Roman space should be less than the time to reject a nonentry (a uniquely Roman letter) in the Cyrillic space. We are reminded again, however, that the opposite result was actually the case.

It is highly questionable, therefore, that the difference in rejection latency is owing to a difference between alphabet spaces in number of representations. Nevertheless, we can preserve the idea that the difference in rejection latency is localized in the mapping from feature description to alphabet space. Consider the presentation of a uniquely Cyrillic letter when the subject is in the Roman mode, that is, when the subject is asked "Is this c an r?" In Tversky's (1977) terms, the target Cyrillic letter (c) is the subject and an individual Roman representation (r), to which it is matched, is the referent. Let s(c,r) be interpreted as the degree to which the subject c is similar to the referent r. We may then take the average latency for rejecting a Cyrillic character as Roman as an index of the degree to which a description of a Cyrillic character is, on the average, similar to a description of a Roman character, that is, as an index of s(c,r). By the same reasoning, the average latency for rejecting a Roman character as Cyrillic may be taken as an index of the degree to which a description of a Roman character is, on the average, similar to a description of a Cyrillic character, that is, as an index of s(r,c). It follows, therefore, that $s(c,r) > s(r,c)$. In words, the descriptions of Cyrillic characters are, on the average, more similar to the descriptions of Roman characters than the descriptions of Roman characters are, on the average, similar to the descriptions of Cyrillic characters.

Asymmetric similarities are not uncommon (see Tversky, 1977) as the use of similes and metaphors readily attests. Thus, we might say that a highway is like a snake, but we would be less likely to say that a snake is like a highway. In this example the snake, noted for winding its way across the ground, is used as the referent rather than the subject of the metaphor. Herein lies a thorny point of theory: the direction of asymmetry depends on which term is the referent. As a general rule Tversky (1977) claims that the determination of subject and referent depends on the relative salience of the objects where the more salient object is assigned the role of referent and the less salient object is assigned the role of subject. Given this, the less salient object is more similar to the salient object than vice versa. In our case, then, we would have to conclude that the representational space of the Roman alphabet is more salient than that of the Cyrillic alphabet. How are we to understand "salient"? Are the set of descriptors (features) for Roman letters more salient--more prominent--than the set of descriptors for Cyrillic letters? It seems reasonable to claim that one set of descriptors is more salient than another if the former includes the latter. However, our intuition, on inspection of Figure 1, is that the set of descriptors for the Cyrillic alphabet includes the set for the Roman alphabet and not vice versa. Experiment VI will provide further reason for doubting a feature-based account of the asymmetry. For the present, we may recognize a less discerning account, namely, that the asymmetric similarity between Cyrillic and Roman is consonant with the view that the device for accepting Roman characters entails the device for accepting Cyrillic characters but not vice versa.

EXPERIMENT II

To assess further the asymmetric similarity between processing Cyrillic letters and processing Roman letters, we consider the phenomenon in the short-term memory literature known as release from proactive interference.

On successive short-term memory tests of the distractor kind (Brown, 1958; Peterson and Peterson, 1959), a subject is given short lists of maybe three items (words, letters, etc.) to retain, with a new list for each test. If the items presented on the successive tasks are drawn from the same category, recall performance across the successive tests will decline precipitously. This is referred to as the build up of proactive interference. If we now present items on a short-term memory test that have been drawn from a category conceptually different from that used in the immediately preceding tests, then there is an abrupt recovery in recall performance. For example, if a subject received four successive tests with digits as the to-be-remembered material and then on the fifth test he was given letters to retain, performance on the fifth test would be similar to that on the first and substantially superior to that on the fourth. In particular, performance on the fifth test would be substantially superior to the recall of the same set of letters after a succession of four tests with letters. Wickens (1970) has proposed that the "release from proactive interference" identifies "psychological" categories. We can assume that there is a common way of encoding within a class (accounting for the decline in recall) that differs between classes (accounting in turn for the increase in recall with shift in class).

We can adopt this strategy to examine the aforementioned asymmetric similarity. By definition, proactive interference is the forgetting induced by earlier items on a later item. The interference is class specific and, ceteris paribus, the more similar the earlier items are to the later item, the greater is the interference and hence the forgetting. Given a succession of five short-term memory tests, we can ask, therefore, how similar the earlier items (those of Tests 1-4) were to the most recent item (that of Test 5). Precisely, we can ask (a) how similar is (the processing and storing of) Cyrillic alphabet material to (the processing and storing of) Roman alphabet material and (b) how similar is (the processing and storing of) Roman alphabet material to (the processing and storing of) Cyrillic alphabet material.

Method

Subjects. The subjects were 360 undergraduate volunteers from the Faculty of Engineering at the University of Belgrade, whose elementary education had been received in Eastern Yugoslavia.

Materials. Ten 8 x 3 inch test cards were prepared, on each of which were printed three letters. Five of the cards contained Cyrillic letters and five contained Roman letters. The five Cyrillic triplets were five different combinations from the letters д, ф, г, л, з; the five Roman triplets were five different combinations from the letters D, F, G, L, Z. These Cyrillic and Roman letter sets are phonemically identical.

Procedure. Each subject received five successive short-term memory tests where each test consisted of the following sequence of events. First, a verbal "ready" signal followed by a letter triplet presented for 3 seconds duration and read aloud by the subject; a three-digit number was then presented from which the subject counted backwards by threes for 10 secs; finally, a recall signal was given with five seconds allotted to recall. A period of 10 seconds elapsed between successive tests.

Design. On appearance at the laboratory, each subject was assigned to one of four groups, with 90 subjects per group. Two groups received letter triplets from the same alphabet on all five tests; thus, one group received only Cyrillic letters for retention and the other only Roman. The remaining two groups were given four successive tests with letters from one alphabet, but on the fifth test were presented letters from the other alphabet. Thus, one group was given four Roman triplets followed by a Cyrillic triplet, and the other was given four Cyrillic triplets followed by a Roman triplet.

Results

The recall of each subject on each test was scored in terms of whether the correct letter was reported in the correct position of a triplet. The averaged results for each condition are given in Figure 6. From inspection of the figure it is evident that proactive interference effects were manifest: performance declines with increasing numbers of short-term memory tests.

The comparisons of interest are these: first, the recall of the Cyrillic triplets on Test 5 after a history of Cyrillic triplets and after a history of Roman triplets; second, and similarly, the recall of the Roman triplets on Test 5 after a history of Roman triplets and after a history of Cyrillic triplets. These comparisons define the release from the proactive interference condition. Precisely, one is interested in whether an item is recalled better from short-term memory when it follows items from a supposed different class than when it follows items from the same class.

The outcome of these comparisons is straightforward. There was most evidently a release from proactive interference ($p < .001$) when the shift was from Roman to Cyrillic (as compared to the all-Cyrillic condition), but hardly a glimmer of release when the shift was from Cyrillic to Roman (as compared to the all-Roman condition).

Before we proceed to entertain this asymmetry seriously, a few cautionary remarks are in order. Only five letters were chosen from each alphabet sample. This, perhaps, obviates the ecological validity of the experiment and introduces the kinds of issues that Clark (1973) has raised about language related experiments. In short, we must be wary of drawing general conclusions about the alphabet distinction on the basis of our limited sampling. Nevertheless, the motivation for using the limited sample should be emphasized: we wished to limit the basis for distinguishing Roman and Cyrillic to visual properties and/or alphabet membership. By the use of two small, phonemically equivalent samples we insured that the transition on Test 5 was not likely one of phonemic content.

Let us, therefore, consider the asymmetry in proactive interference. To reiterate, the release from the proactive interference paradigm is essentially an experimental embodiment of the question: How similar is the processing of x to the processing of y? In the present case, x is the alphabetic material presented on Tests 1-4 and y is the alphabetic material presented on Test 5. We can therefore identify x with the subject of the similarity comparison and y with the referent. In that the shift from Roman letters on Tests 1-4 to Cyrillic letters on Test 5 yielded a release from proactive interference, we may claim that processing Roman letters _is_ _not_ _very_ _similar_ to processing Cyrillic letters. In that a shift from Cyrillic letters on Tests 1-4 to Roman letters on Test 5 yielded no release from proactive interference, we may claim that processing Cyrillic letters _is_ _very_ _similar_ to processing Roman letters. [As in Experiment I, $s(c,r) > s(r,c)$.]

Finally, before leaving this experiment, we should note that the degree of proactive interference in the first four Cyrillic tests was substantially greater than in the first four Roman tests, suggesting that the Cyrillic letters used were more visually confusable than the Roman.

EXPERIMENT III

As remarked, the first experiment does not mimic any especially natural situation. The Yugoslavian is rarely called upon to explicitly label the alphabet in which he is reading; the alphabet, by all accounts, is transparent to the reading process. However, a circumstance in which the Yugoslavian, particularly the Eastern Yugoslavian, often finds himself is one in which he must flit back and forth between the two alphabets as he reads posters, street signs, shop names and the like. In the cities the two alphabets are used with abandon. We may suppose, therefore, that in order to keep the ambiguous letters straight, the local inhabitant must detect the structure of the letter string that specifies whether the word is a Cyrillic word or a Roman word. In short, there ought to be a means by which he can rapidly determine the alphabet without having to identify the letters. In the present experiment and the one that follows, we are interested in demonstrating that the Serbo-Croatian reader has this facility, precisely, to determine alphabet before determining identity. However, first let us make some preliminary, but necessary, remarks on the research that is the backdrop for our third and fourth experiments.

How does one detect the presence or absence of a specified letter in an array of letters? At first blush we might conjecture that, in principle, visual search is a matter of hunting for the right combination of visual features. This point of view, espoused by Neisser (1967), has received considerable support from the reliable observation that search times relate inversely to the visual similarity between the target letter and the foil letters. Phonetic similarity between target and foils proves to be a far less significant determinant of search performance. So we might suppose, after Neisser (1967), that in the letter search situation, visual feature analyzers irrelevant to the target can be turned off; in Broadbent's (1971) terms, searching for a given letter is a matter of "filter-setting."

Unfortunately, this treatment of the letter-search process is rudely shaken by the observation that category distinctions between the target and background items are not immaterial to the search. Posner (1970), Brand (1971), Ingling (1972) and Egeth, Jonides and Wall (1972) have all demonstrated that when looking for a specified character, latency of search is significantly shorter when the target is embedded in an array of characters from another category. Thus, one can search for a letter (digit) faster when the foils are digits (letters) than when the foils are letters (digits). Also we should note that a comparable result is obtained in paradigms that are not strictly identical to the visual search procedure (for example, Sperling, Budiansky, Spivak and Johnson, 1971).

Of course, one could argue that the above "category effect" is due to the fact that letters as a set and digits as a set are visually distinguishable; particularly features are more prevelant in one set than in the other. Two experiments, however, militate against this argument. In one (Ingling, 1972), the other category foils were chosen to be as similar as possible to the target--a manipulation, however, that did not eliminate the category effect. In the other (Jonides and Gleitman, 1972), the ambiguous character 0 was identified prior to search as "Ō" or as "zero." The latency of search for the 0 was determined by the relation between how it was identified and the class of the foils, for example, searching for 0 in an array of letters was faster when 0 was conceptualized as "zero" rather than as "Ō." We may refer to this phenomenon as the "conceptual category effect." At all events, it would appear that, if conditions permit, searching for a given character can be governed by category or pigeon-hole setting (Broadbent, 1971) rather than by filter setting.

Our third and fourth experiments are, essentially, Roman/Cyrillic analogues of the aforementioned letter/digit experiments. Thus, the third experiment asks whether searching for a letter in an array of letters from the other alphabet is faster than searching for a letter in an array of letters from the same alphabet.

Method

Subjects. The subjects were 26 undergraduate students from the Faculty of Engineering, University of Belgrade. They had received their elementary education in Eastern Yugoslavia. Each subject was paid the equivalent of $2.00 per session.

Materials. The letters were Letraset black uppercase (Helvetica Light, 12 points). Sixteen letters, eight uniquely Roman and eight uniquely Cyril-

Serbo-Croatian

Roman		Cyrillic		
Printed Upper Case	Cursive Lower Case	Printed Upper Case	Cursive Lower Case	Letter Name in I.P.A.
A	a	А	а	a
B	b	Б	б	bə
C	c	Ц	ц	tsə
Č	č	Ч	ч	tʃə
Ć	ć	Ћ	ћ	tʃjə
D	d	Д	д	də
Đ	đ	Ђ	ђ	dʒjə
DŽ	dž	Џ	џ	dʒə
E	e	Е	е	ɛ
F	f	Ф	ф	fə
G	g	Г	г	gə
H	h	Х	х	xə
I	i	И	и	i
J	j	Ј	ј	jə
K	k	К	к	kə
L	l	Л	л	lə
LJ	lj	Љ	љ	ljə
M	m	М	м	mə
N	n	Н	н	ne
NJ	nj	Њ	њ	njə
O	o	О	о	ɔ
P	p	П	п	pə
R	r	Р	р	rə
S	s	С	с	sə
Š	š	Ш	ш	ʃə
T	t	Т	т	tə
U	u	У	у	u
V	v	В	в	və
Z	z	З	з	zə
Ž	ž	Ж	ж	ʒə

TABLE 1

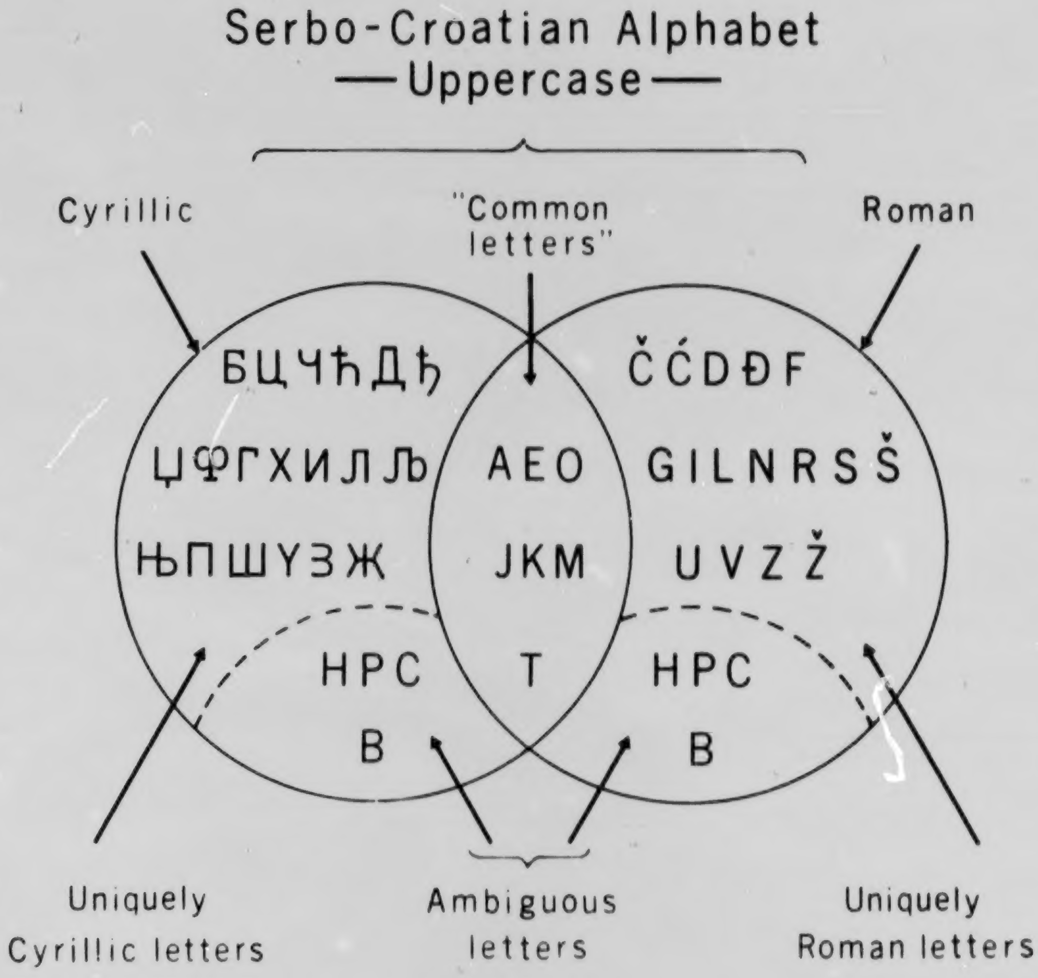

Serbo-Croatian Alphabet
—Uppercase—
Cyrillic
"Common letters"
Roman
БЦЧЋДЂ
ЏФГХИЛЉ
ЊПШУЗЖ
AEO
JKM
T
ČĆDĐF
GILNRSŠ
UVZŽ
HPC
B
HPC
B
Uniquely Cyrillic letters
Ambiguous letters
Uniquely Roman letters

Figure 1: The Roman and Cyrillic alphabets (uppercase).

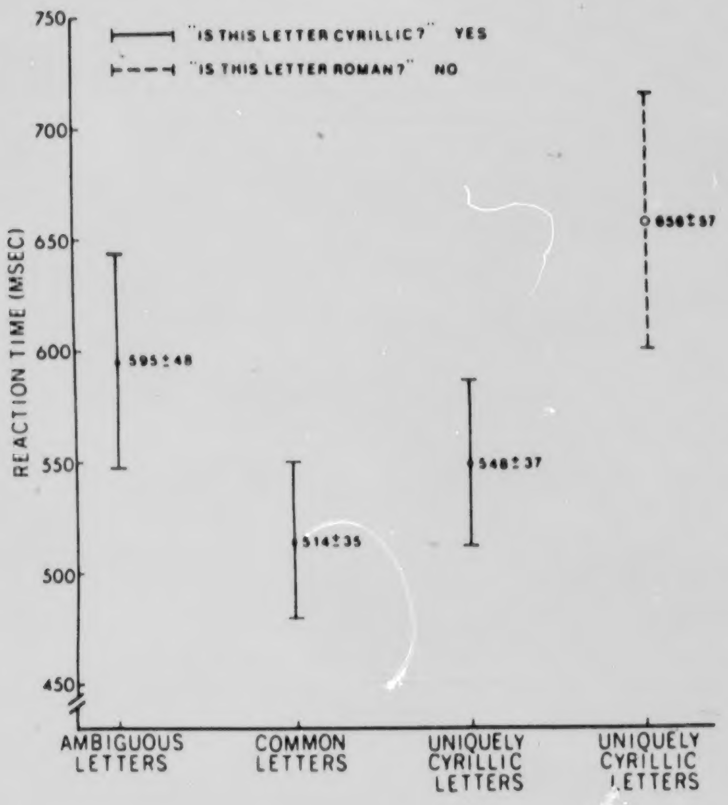


Figure 2: Alphabet decision latency in Experiment I.

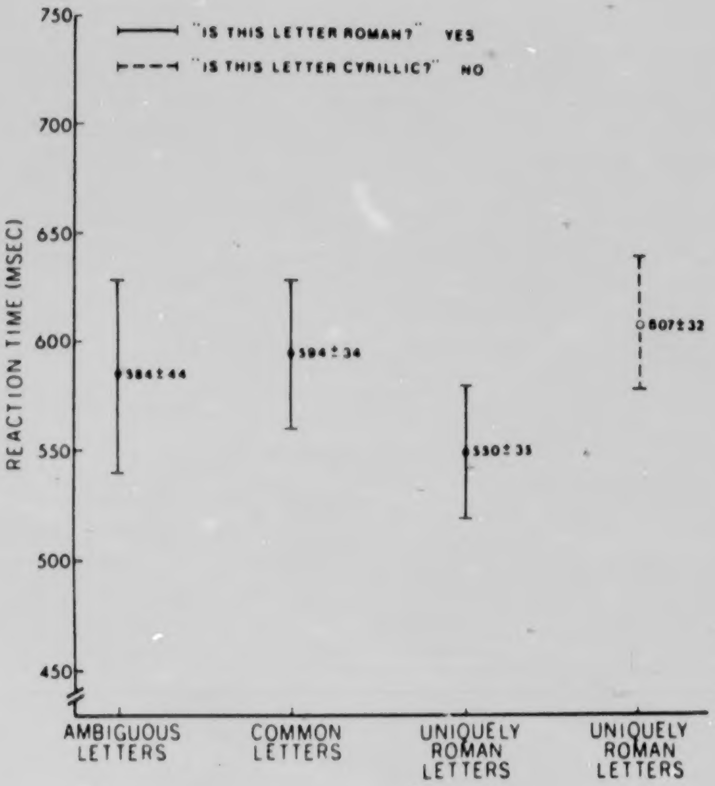


Figure 3: Alphabet decision latency in Experiment 1.

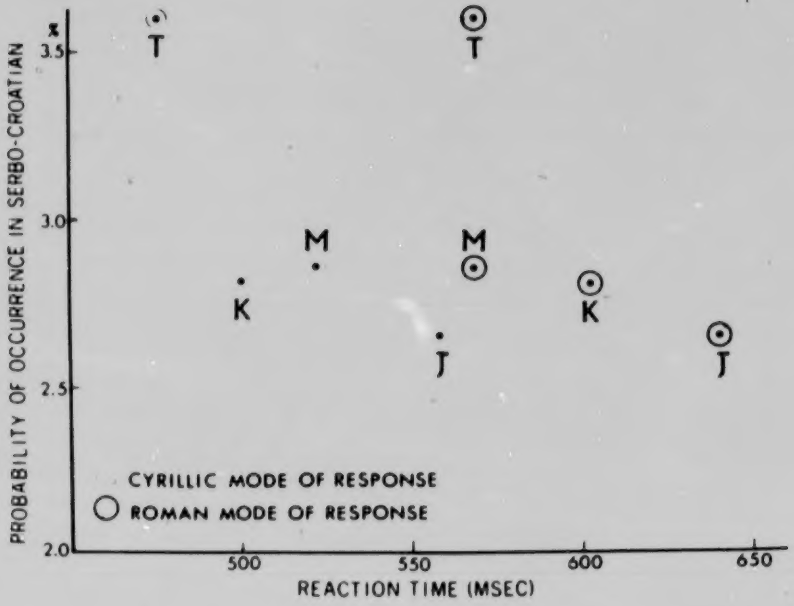


Figure 4: Relation between probability of occurrence and latency of alphabet decision for the common letters M, K, T, J.

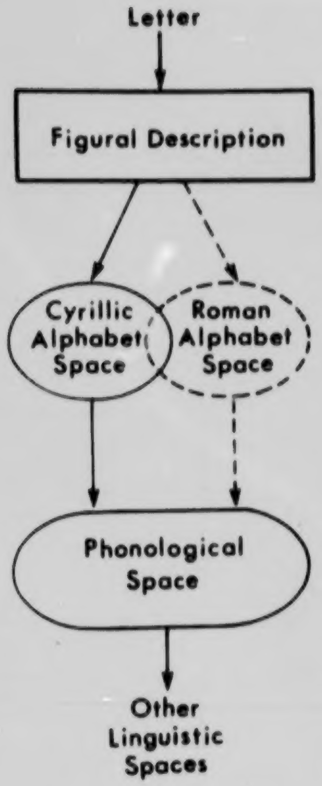

Letter
Figural Description
Cyrillic Alphabet Space
Roman Alphabet Space
Phonological Space
Other Linguistic Spaces

Figure 5: Possible stages in the processing of Serbo-Croatian characters and the relation between the first- (Cyrillic) and second-learned (Roman) alphabet.

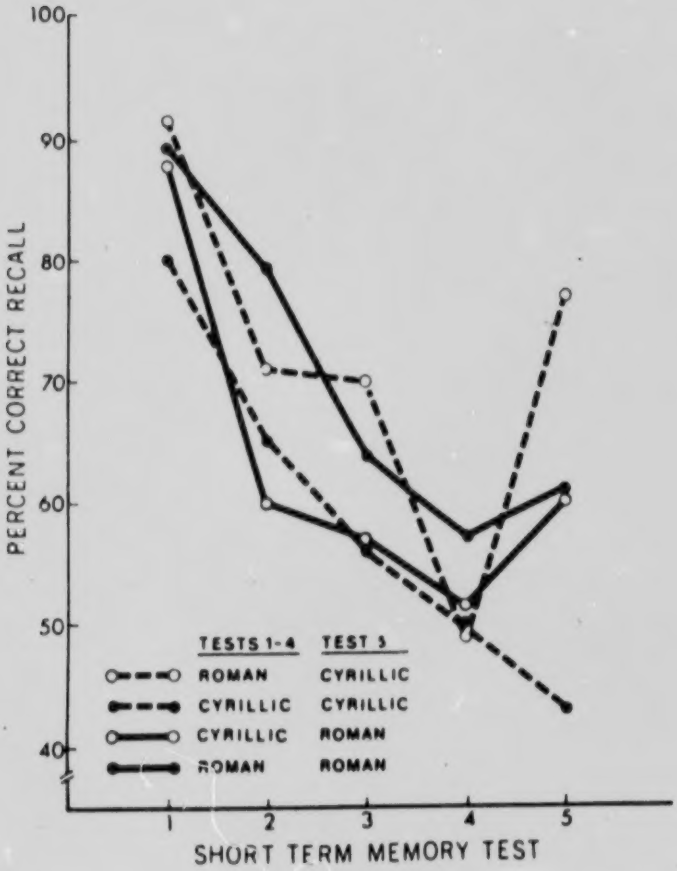


Figure 6: Recall as a function of short-term memory test in Experiment II.

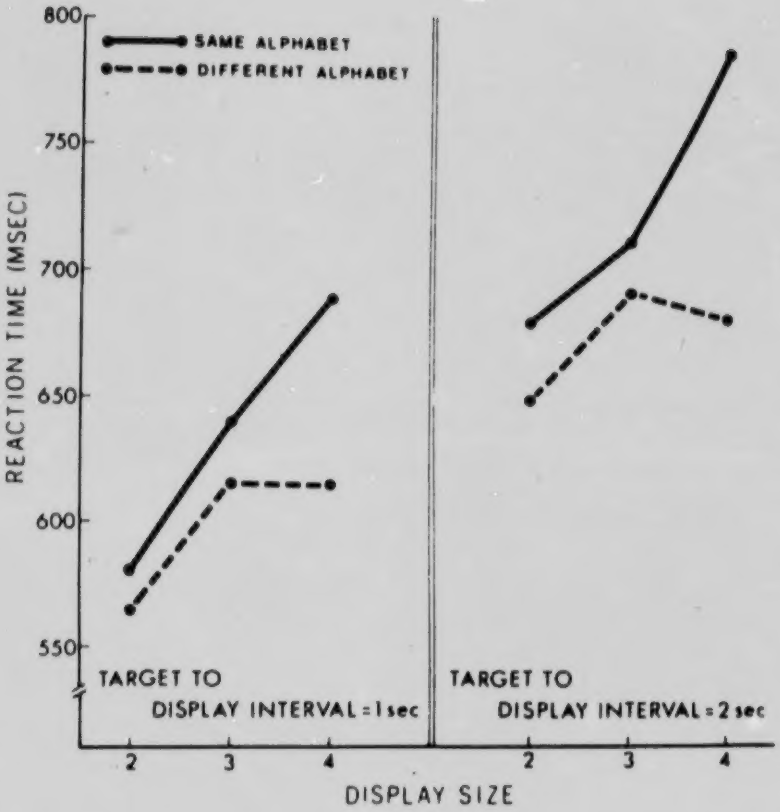


Figure 7: Reaction time as a function of display size at two target-to-display intervals in Experiment III.

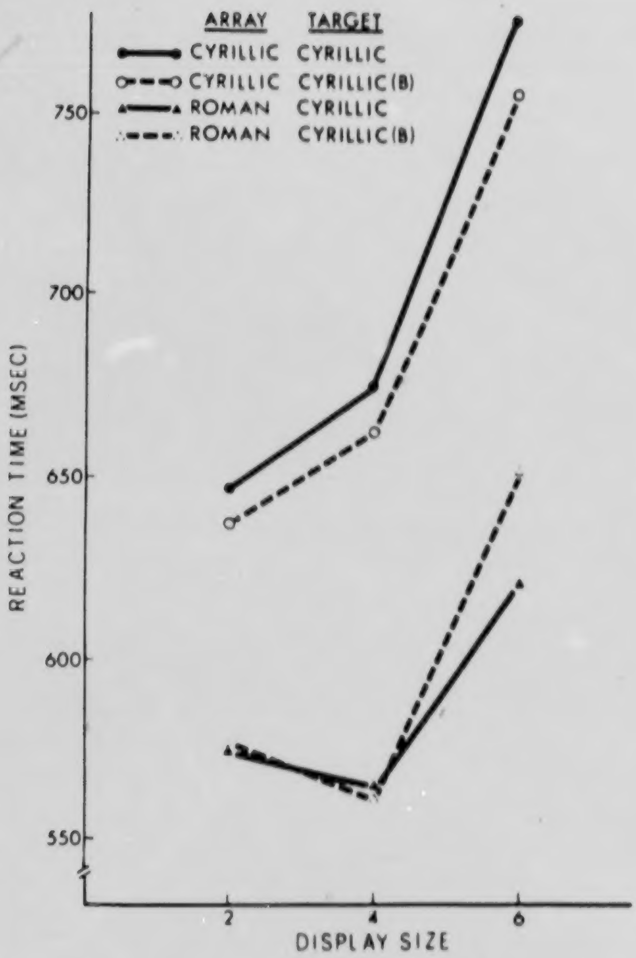


Figure 8: Latency of searching for Cyrillic targets as a function of display size in Experiment IV.

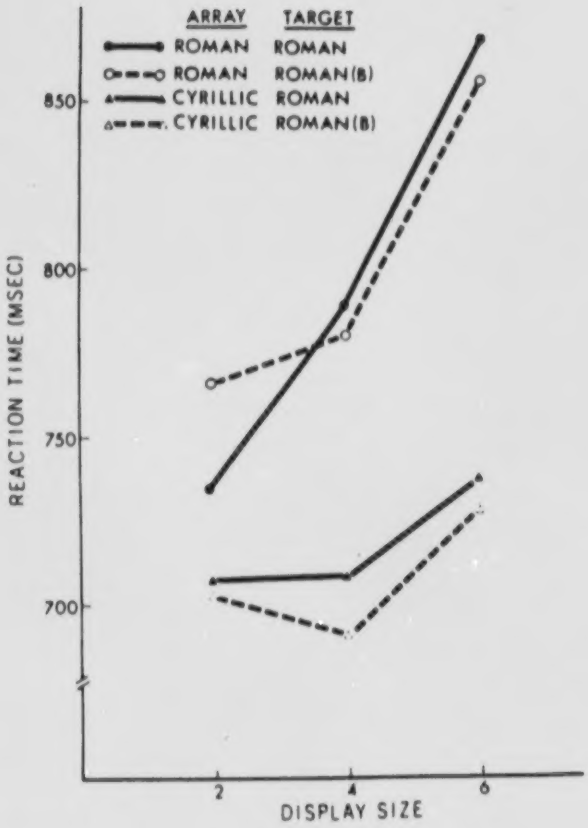


Figure 9: Latency of searching for Roman targets as a function of display size in Experiment IV.

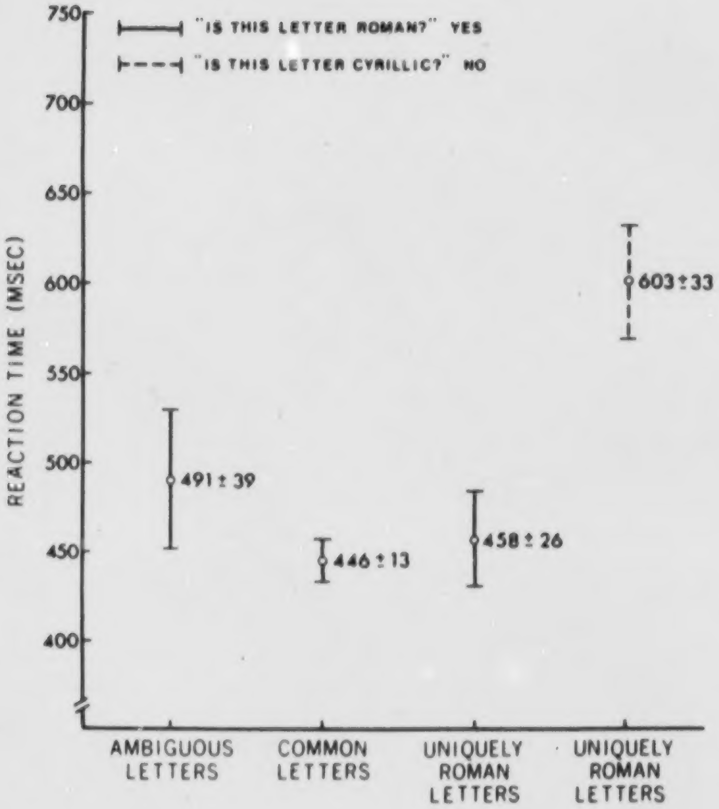


Figure 10: Alphabet decision latency in Experiment VI.

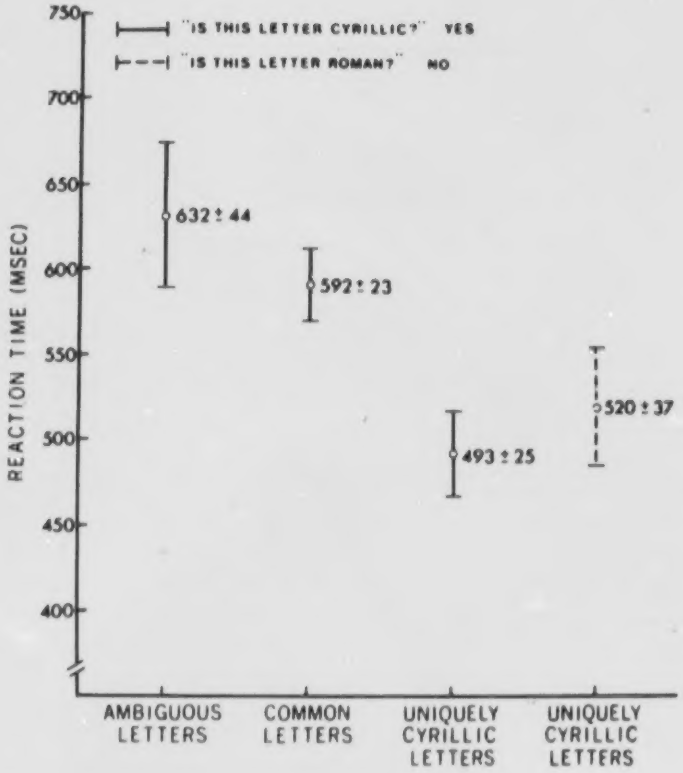


Figure 11: Alphabet decision latency in Experiment VI.

lic, were used to construct 100 pairs of target and array slides. The search field arrays were quasi-randomly constructed (through a Latin square design) from the Roman or the Cyrillic letters. The items in a search field were 2, 3, or 4 in number, and they were located around the circumference of an imaginary circle whose center coincided with a preexposure fixation point. To keep overall visual angle constant, the following injunctions were met: when there were only two items, they were located on a slide in diametrically opposed locations on the imaginary circle; when there were three or four letters, two were located in diametrical opposition and the others located randomly (see Egeth et al., 1972; Jonides and Gleitman, 1972). The letters in the set of target displays were centered so as to overlay the preexposure fixation point. The three channels of a Scientific Prototype tachistoscope were used to present the exposures.

Design. The subjects were assigned, on order of appearance at the laboratory, to one of two groups with 13 subjects per group. The two groups were distinguished by the interval elapsing between the target exposure and the search field. For one group this interval was one second, for the other it was two seconds. In each group the target's relation to the search array--same or different alphabet--was combined factorially with two response types (positive and negative) and three levels of array size (2, 3, or 4). More precisely, the two response types were whether or not the target was in the search field.

Procedure. A trial consisted of the following events: an auditory warning signal followed immediately by a target field (single letter) exposure of one second followed, in turn, one or two seconds later by a search field of 200 msec duration. The preexposure, target and search fields were 10.3 cd/m^2. The onset of a search field triggered an electronic timer that was stopped when the subject pressed either the "yes" key to indicate target presence or the "no" key to indicate target absence. The key-press technique was the same as that described for Experiment I.

Fifty practice trials were followed by 150 trials organized with brief rest periods consequent to every 25. "Yes" and "no" responses were equally distributed across the 150 trials.

Results and Discussion

Within the two groups, that is, the one-second, target-to-array interval group and the two-second, target-to-array interval group, the same alphabet and the different alphabet conditions were compared. Mean reaction times were computed for each subject in each condition at each search field size, ignoring errors that occurred at a mean rate of 2.5 percent. For simplicity, only negative responses are considered, that is, responses for the trials when the target was not present in the search field.

Figure 7 plots the contrast between searching for a target in an array of letters from the same alphabet and searching for a target in an array of letters from the other alphabet. For both intervals different-alphabet search is obviously faster than same-alphabet search ($F = 32.96$, $df = 1, 24$; $p < .01$) in keeping with the comparable contrast in the letter/digit search experiments. In brief, this experiment corroborates the thesis that, where conditions permit, one can encode alphanumeric materials categorically prior to more complete identification, and that such encoding can facilitate the processing rate in search tasks (Ingling, 1972). We will reserve further comment on this issue until the discussion of the fourth, and related, experiment. For the present, we address the less significant issue of why the latencies were slower ($F = 7.80$, df = 1, 24; $p < .01$) for the longer target-to-array interval. Inasmuch as Figure 7 and the analysis of variance ($F < 1$) gives us no reason for believing that the slopes of the functions differed from one interval to the next, it would seem that the most appropriate interpretation would be one having to do with the status of the target representation. Let us explain. The slope may be taken as indicative of comparison time; our data are for negative responses, so we can legitimately assume exhaustive search (and we recognize that there are both serial and parallel search models that would accommodate those functions). The difference between the one-second to the two-second condition is the intercept. Now if the representation of the target is changing over the interval between the target and the search array (for example, Posner, 1969), then we might assume that the extra intercept time in the two-second condition reflects operations on the target representation. The goal of these operations might be that of putting the target representation into a form permitting visual/alphabetic comparison. Presumably such operations, if needed, were less time consuming in the one-second condition.

EXPERIMENT IV

The fourth experiment departs only slightly from the third. Purportedly it aimed at maximizing the alphabet differentiation ability demonstrated in Experiment III. To this end the participants were instructed that, given a target from one of the alphabets, there would never be a case in which that target would occur in a search array of letters from the other alphabet. In other words, when the search array was presented, the participant was encouraged by the instruction to first determine the alphabet, for by so doing he could save himself the trouble of searching for the target on those trials in which the alphabet of the array differed from that of the target.

There was one further major difference between Experiments III and IV. The present experiment used an ambiguous letter--B--in the target set. Following Jonides and Gleitman's (1972) example with 0, one group of subjects was told that B was Roman, another than it was Cyrillic. Would search performance with B be comparable to that with a nonambiguous letter?

Method

Subjects. The subjects were 34 undergraduates from the same pool as that used in Experiment III. Each was paid the equivalent of $2.00 per session.

Materials. A total of 19 letters were used to prepare the target and search fields. Nine of these were uniquely Cyrillic, nine were uniquely Roman and one was the ambiguous letter B. Target and search fields were constructed in the fashion described in Experiment III, except that the sizes of the search array were 2, 4, and 6 letters.

Design. Each subject was assigned to one of two groups by order of appearance at the laboratory. There were 16 subjects in one group, 18 in the other. One group was designated Roman; they were told at the outset that their targets were Roman and would be throughout the experiment. They were informed of the three targets: D, F, and B. The other group was designated Cyrillic; they were told at the outset that their targets, for the duration of the experiment, were Cyrillic: Д , ф and B. For both groups there were simply three target/search field relations: (1) a target was present in a search field of the same alphabet; (2) a target was not present in a search field of the same alphabet; (3) a target was not present in a search field of the other alphabet.

Procedure. A trial was defined as in Experiment III. The target-to-array interval was two seconds. There was a total of 150 trials with an equal number of positive and negative responses.

Results. For all reaction time analysis, only the negative responses are considered and data from error trials (approximately 4.5 percent) were excluded.

The mean reaction time at each array size for each subject in each condition of the Roman group was entered into an analysis of variance. The Cyrillic data were similarly organized and entered into a separate analysis. Both analyses were within-subject, repeated measures. In both the Roman and Cyrillic cases there was a significant effect of target-to-array alphabet relation (same or different): $F = 12.35$, df = 1, 90, $p < .001$ and $F = 16.36$, df = 1, 102, $p < .01$, respectively. Similarly, in both cases, array size was a significant variable: $F = 4.51$, df = 2, 90, $p < .05$, and $F = 4.01$, df = 2, 102, $p < .05$, respectively. The Roman and Cyrillic group data are displayed in Figures 8 and 9. The figures also give the corresponding functions for the ambiguous target, B. As can be seen, the B functions in the Roman and Cyrillic cases do not differ from those of the uniquely Roman or uniquely Cyrillic targets.

Discussion

The third and fourth experiments provide unequivocal evidence that the Yugoslavian reader of two alphabets can readily distinguish the visual appearance of one alphabet from that of the other and that alphabet classification could well anticipate letter and, in consequence, word recognition. As we have remarked before, it would be to the benefit of the Yugoslavian, in view of the presence of ambiguous letters, to have at his disposal a means of rapidly determining the alphabet in which a word is written.

The most parsimonious explanation of the data of these two experiments is that there is a general physical difference between the uniquely Cyrillic and the uniquely Roman letters. This is an intuitively sound explanation as the reader can verify for him or herself by examining Figure 1. Nevertheless, as we noted in the introduction of Experiment III, those who have observed the "category effect" with respect to the letter/digit distinction have not been so willing to assume that it is owing simply to some, as yet undefined, physical difference. In a way, we can sympathize with this reticence; after all, it is not obvious what physical differences might separate letters from digits as a class. There is, in addition, the quite remarkable discovery of Jonides and Gleitman that the category effect in visual search can be obtained when the target is conceptually rather than physically defined. The upshot of their experiment, we recall, is that the category effect is not an artifact of a simple physical difference between the target and the background items.

We are forced, therefore, to accept with caution the claim that search in the different alphabet conditions of Experiments III and IV was faster than in the same alphabet conditions because of a physical contrast. Perhaps the distinction is more abstract. Unfortunately our fourth experiment, although it uses the ambiguous character B, does not simulate the design that permitted Jonides and Gleitman to draw their unequivocal conclusion. The design had subjects search through arrays, knowing full well that regardless of the array's alphabetic relation to the target, the target had a good chance of being present. In short, Jonides and Gleitman's subjects had to search; our subjects did not.

The fact remains that our data and those of Jonides and Gleitman are very similar; further, the difference at array size 4 in our Experiment III is comparable to that at array size 4 in our Experiment IV. In Experiment III, the subjects had to search. So, perhaps, we are mistaken in assuming that the subjects in Experiment IV behaved differently from those in Experiment III. In sum, perhaps the result we obtained with the ambiguous letter B in the fourth experiment is the same as the result Jonides and Gleitman report; and, further, that it is owing to the same reason, namely, a conceptual rather than a figural difference between one class and the other.

Let us conclude this discussion by noting that overall performance in Experiment IV was substantially superior, that is, latencies were lower for Cyrillic search arrays than for Roman search arrays. The latency difference is not due to differences in rate of search per se. In the Roman case the slope for the same alphabet condition was 33 msec/letter, and for the different alphabet condition it was 7.9 msec/letter. The corresponding values in the Cyrillic case were 32.5 and 11.5. The difference between the two alphabets in this regard is found at the intercept value: that for the Cyrillic case is, on averaging, 555.5 msec and that for the Roman, 677 msec. If our subjects are differentiating alphabet antecedents to determine identity, then, apparently, the Cyrillic is distinguished more rapidly than the Roman.

EXPERIMENT V

If, in the temporal course of information processing, a distinction can be drawn rapidly between the two alphabets, we may inquire as to the first stage at which the distinction is manifest. Given one popular view of the flow of visual information (for example, Neisser, 1967; Haber, 1969), the first significant stage is the transient medium of literal storage referred to as the icon (Neisser, 1967). However, the general consensus is that at the level of iconic storage, derived distinctions--symbolic distinctions--are not made (for example, Coltheart, 1975). There is ample evidence that selection from iconic storage can proceed efficiently when the criterion for selection is some physical property such as size, color, location, etc., but that it is extremely poor when the criterion is category (for example, letters or digits). The conventional wisdom favors the view of the icon as precategorical (Dick, 1974). However, if the Roman/Cyrillic distinction is founded on a less abstract contrast than that which permits the differentiation of letters and digits, that is, that the two alphabets are distinguishable by general physical properties, then it might prove to be the case that iconic memory is the first stage at which the alphabet distinction arises. Experiment V was designed to test this possibility.

The technique used was delayed partial-sampling (Sperling, 1960). The observer is presented an array of letters (in the present experiment the array is arranged as two rows of four or three rows of three) exposed very briefly, and the observer's task is to report either as many letters as he can (whole report) or a subset of the total number of letters (partial report). In the latter case, the subset to be reported is specified by a signal given after the exposure has terminated. Generally, the partial report, as an estimate of the number of items available to the observer subsequent to the exposure, exceeds that of whole report. However, it is argued that this superiority will hold if and only if the basis for partial report (the selection criterion) has been differentiated at the level of processing that supports the persistence of the array beyond its exposure. In short, whether or not partial report by alphabet is superior to whole report will depend on whether or not this alphabet distinction actually exists at the level of iconic persistence. The foregoing, for all intents and purposes, defines the logic of Experiment V.

Method

Subjects. Thirty students from the same population used in the previous two experiments served as subjects. They received the equivalent of $2.00 per session.

Materials. The two array patterns were 2 by 4 and 3 by 3. Mixed arrays were constructed from a set of nine uniquely Roman and nine uniquely Cyrillic letters. For the construction of pure arrays--that is, arrays that were of one alphabet--three extra letters were used. These were the ambiguous letters C, H, B. A total of 72 mixed and 72 pure arrays were constructed from black uppercase letters (Helvetica Light, 12 points). In all arrays, a letter appeared in each of the possible positions equally often. This meant that the dispersion of Roman and Cyrillic letters in a mixed array was haphazard.

Presentation of Displays. The array exposure duration was 30 msec. Each array was preceded and followed by a fixation field containing a black fixation point at its center. The array and fixation field were 10.3 cd/m^2 and were projected in two channels of a Scientific Prototype three-channel tachistoscope. For the partial report situation, the subject was equipped with earphones and received one of the two tones simultaneous with the offset of the array. A high tone (3000 Hz) signaled the report of one alphabet, a low tone (300 Hz) signaled the report of the other. The relation between array and tone was determined in a quasi-random fashion.

Procedure. The subject was instructed to look at the fixation point and, when ready, to press a button with a finger of his left hand. This triggered an auxiliary electronic unit which in turn, after a 500 msec delay, initiated the exposure of an array.

In both whole and partial report conditions, the subject recorded his responses on a response grid in which the cells corresponded to locations of the array. For each trial a new response grid was used; the subject, therefore, did not have visual access to his prior responses. For the whole report, the subject was instructed to write down as many letters in their correct locations as he could read, guessing when he was not certain. For partial report, the subject was required to report only the letters from the alphabet signaled by the tone.

Design. The whole session of 144 trials was divided into four blocks. The first block consisted of 36 pure arrays; the second and third blocks each consisted of 36 mixed arrays; and the fourth and final block was again 36 pure arrays. Within each block there were 18 4/4 arrays and 18 3/3/3 arrays.

The subjects were divided into two groups as a function of the alphabet that made up the pure arrays of the first and fourth blocks. For one group this alphabet was Roman; for the other group the alphabet was Cyrillic. For both groups, blocks two and three were the same with the tone-alphabet relation counterbalanced across the two groups.

The whole report data were collected from the pure arrays and from the mixed arrays. The latter estimate, however, was collected in an experimental session separate from that described above.

Results and Discussion

Response grids were scored in terms of correct letters reported in their correct positions. Averaging the data over array arrangements revealed that whole report for pure Roman arrays was 3.5 letters and for pure Cyrillic, only 2.8 letters. In Experiment II we had noticed that proactive interference was more pronounced for Cyrillic letters than for Roman. Taken together, Experiment II and the present experiment suggest that the distinctiveness of Cyrillic letters is not as optimal as that of Roman. In a phrase, Cyrillic letters are more likely to confuse with Cyrillic letters than Roman letters are likely to confuse with Roman letters.

If we take the average of the two pure-alphabet whole reports, then we have a value of 3.15 letters; this is equivalent to the whole report estimate from mixed arrays, which was 3.10 letters.

When subjects were required to give partial report, the average number of letters reported from the mixed 2 by 4 arrays was 1.59, and from the mixed 3 by 3 arrays, 1.42. To obtain the estimate of letters actually available to the observer, we follow the general logic of Sperling (1960) and multiply the number of letters reported from the cued subset by the number of subsets. The argument is that if the observer could report x items from a subset cued after the exposure, and if there are y subsets, then the observer must have had in memory xy letters. By this argument, we calculate that the number of available items under conditions of partial report averaged over the two array arrangements is 2.95, and the question to which the experiment was directed is now answered: when delayed partial-sampling is based on the distinction between Roman and Cyrillic letters, partial report is not superior to whole report. In short, we can infer that the distinction between the alphabets is not made at the level of iconic storage.

EXPERIMENT VI

The sixth experiment focuses on the asymmetric relation between processing Cyrillic and processing Roman characters. The fundamental conclusion of Experiments I and II was that processing Cyrillic characters was more similar to processing Roman characters than vice versa. In notation, this asymmetric similarity was expressed $s(c,r) > s(r,c)$. Following Tversky's (1977) argument, however, $s(c,r) > s(r,c)$ iff $f(R) > f(C)$; that is to say, processing Cyrillic is more similar to processing Roman than vice versa, if and only if processing Roman is overall more salient than processing Cyrillic. The problem with defining salience in the present context was remarked upon in the discussion of Experiment I. If, as was presumed in that discussion, the asymmetric similarity arises in the mapping from a character's feature description to the alphabet spaces (see Figure 5), then the salience of the Roman alphabet processing might be interpreted in terms of features. For example, we might say that the dimensions of description of the Roman alphabet include those of the Cyrillic; or that the descriptors of the Roman alphabet distinguish Roman characters more efficiently than the descriptors of the Cyrillic alphabet distinguish Cyrillic characters.

At all events, salience in the preceding is defined as an absolute property of the set of alphabet characters. If true, the direction of asymmetry should be indifferent to the order in which the alphabets are acquired. An alternative view was expressed at the outset of this paper, namely, the device developed for accepting characters of the alphabet acquired second necessarily entails the device for accepting the characters of the alphabet acquired first. On this view, the direction of asymmetry should be very sensitive to the order in which the alphabets were acquired. Precisely, if we replicated Experiment I with subjects who had acquired Roman first and Cyrillic second, then the pattern of results represented in Figures 2 and 3 should be reversed. Experiment VI is such a replication.

Method

Subjects. Twenty-eight subjects were recruited from the Department of Psychology at the University of Belgrade. These subjects had received their elementary education in Western Yugoslavia. They all had considerable experience in reaction time experiments.

Materials and Design. The same letters as used in Experiment I served as the stimulus materials for the sixth experiment. One exception was that the Cyrillic letter X was excluded.

The design of the experiment followed that detailed in Experiment I. The twenty-eight subjects were divided by order of appearance at the laboratory into two groups of fourteen each. One group was instructed to respond to the question "Is this letter Roman?" and the other was instructed to respond to the question "Is this letter Cyrillic?" Each subject saw and responded to 144 slides with each letter appearing a minimum of three times.

Results

Only correct responses were analyzed. The error rates in accepting uniquely Roman letters as Roman and uniquely Cyrillic letters as Cyrillic were, respectively, 3.3 percent and 4.5 percent. The error rate in rejecting uniquely Cyrillic letters as Roman was 7.5 percent and that in rejecting uniquely Roman letters as Cyrillic was 9.0 percent. For common letters, the Roman mode yielded 2.2 percent errors and the Cyrillic mode yielded 6.8 percent errors. For ambiguous letters, the Roman mode yielded 6.2 percent errors and the Cyrillic mode yielded 21.0 percent errors.

The mean reaction times for each letter within a class were averaged across subjects and then the class average was determined. The results are shown in Figures 10 and 11.

As with Experiment I, the subjects behaved differently under the two question regimes. However, as comparison of Figures 10 and 11 with Figures 2 and 3 dramatically reveals, under the two question regimes, the behavior of the subjects indigenous to Western Yugoslavia is diametrically opposite to that of the subjects indigenous to eastern Yugoslavia. For the subjects of the present experiment, the common letters were accepted as Roman letters much more rapidly than they were accepted as Cyrillic letters ($t = 10.79$, $df = 22$, $p < .001$). The converse was found to be true in Experiment I. The present experiment, like the first, reveals little difference between the two question regimes when the class of letters is unique and the response is "yes," but a substantial difference between the two regimes for the unique letters when the response is "no." However, the difference is in the opposite direction to that of Experiment I, that is, the subjects of the present experiment found it much more difficult to reject a Roman letter as Cyrillic than to reject a Cyrillic letter as Roman ($t = 7.20$, $df = 22$, $p < .001$).

Finally, we can consider the ambiguous characters. In the first experiment the latency for accepting the ambiguous letters as Roman was approximately the same as the latency for accepting them as Cyrillic; and in both question regimes, these acceptance latencies were slower than for the unique characters. For the present experiment it remains the case that ambiguous characters are accepted more slowly than the uniquely Roman and the uniquely Cyrillic characters ($t = 2.81$, $df = 13$, $p < .05$ and $t = 9.75$, $df = 13$, $p < .001$), although an analysis of latencies cannot be taken too seriously in view of the error rate. Nevertheless, inspection of Figures 10 and 11 and a consideration of the error rates leads to the conclusion that the subjects of the present experiment found it much more difficult to accept the ambiguous letters as Cyrillic than as Roman.

Discussion

We concluded in the discussion of Experiment I that the subjects viewed the common letters as essentially members of the Cyrillic alphabet and only indirectly as members of the Roman alphabet. That conclusion for Eastern Yugoslavian subjects most obviously does not hold for the Western Yugoslavian subjects of the present experiment. For the latter we would have to concede the common letters to the Roman alphabet space and only indirectly to the Cyrillic. Clearly, the allegiance of the common letters to one or the other alphabet is determined by which alphabet is learned first.

It is also clear, in the contrast of the present experiment with the first, that the asymmetric similarity between Roman and Cyrillic processing is tied to the order in which the alphabets are learned and not to any absolute structural difference between the two alphabets. The finding of the first experiment, that rejecting Cyrillic letters in the Roman mode takes longer than rejecting Roman letters in the Cyrillic mode, led us to the understanding that, in some sense and at some level, processing Cyrillic is more similar to processing Roman than vice versa. A comparable consideration of the rejection latencies of the present experiment, however, leads to the opposite asymmetry: in some sense, and at some level, processing Roman is more similar to processing Cyrillic than vice versa. In notation, the asymmetry for the subjects indigenous to Western Yugoslavia is $s(r,c) > s(c,r)$; for subjects indigenous to Eastern Yugoslavia it is, as noted above, $s(c,r) > s(r,c)$.

CONCLUSION

The secondary findings of the present experiments can be summarized briefly, indicating that the (Eastern) Yugoslavian readily distinguishes between the Roman and Cyrillic alphabets and, in principle, could do so prefatory to reading (Experiments III and IV), and that the distinguishing of the alphabets occurs at some information-processing stage subsequent to iconic memory (Experiment V).

The primary finding can similarly be summarized: for a person who has learned the Cyrillic (Roman) alphabet first, there is a sense of processing in which it can be said that processing the Cyrillic (Roman) characters is more similar to processing the Roman (Cyrillic) characters than vice versa (Experiments I, II and VI). We interpret the processing asymmetry and the dependence of its direction on the order of acquisition by saying that whatever the means by which a person has come to read the first-acquired alphabet, those means are adopted to the task of reading the second-acquired alphabet. More precisely, the mechanism for processing the second-acquired alphabet entails the mechanism for processing the first-acquired alphabet, but not vice versa.

The proposed relation between the two alphabets is, perhaps, not dissimilar to the relation between speech and reading, on the one hand, nor on the other hand, to the relation between two languages (bilingualism). One popular, abstract treatment of the acquisition of reading goes as follows: suppose that you had at your disposal a mechanism for understanding language by ear and that your task was to construct a mechanism for understanding language by eye. A wise strategy would be to build an addendum to the available language understander that converted the optical information into a form consistent with the language understander and did so at the earliest possible (reasonable?) level of processing. Given this strategy, it would follow that the mechanism for language by eye necessarily entails the mechanism for language by ear, but not vice versa.

The description of the mechanism for bilingualism is often cited in two roughly distinguished forms (see Reynolds and Flagg, 1977). In one form (the coordinate view), it is contended that the computational support for one language is largely separate from that of the other, even to the extent that the semantic spaces are separate. In the other form (the compound view), the two languages share processing components; in particular they have a common semantic space. Our investigations into bi-alphabetism have assumed, at the outset, a common phonologic space. The claim of a common semantic space for bilingualism is contingent, in part, on the manner in which the languages were acquired. If they were acquired in the same setting or if the learning of the second language was parasitic on the first, then it can be assumed that identical semantic values are ascribed to the corresponding lexical entries and phrase structures of the two languages, resulting in a single, common semantic space. Where the cultural and environmental settings of the learning of the languages differ, then the assumption of a common semantic space is less appealing. This crude and largely inadequate (see Reynolds and Flagg, 1977) differentiation of conditions of bilingual acquisition is of relevance to the Serbo-Croatian bi-alphabetism. Since the setting is invariant for the two alphabets, and since the second alphabet is acquired through the medium of the first, then the phonologic space should not differ between the two alphabets.

There is a sense, then, in which the bi-alphabetism investigated in the present paper relates to the issues of second language learning and the interrelation of a bilingual's two languages. In both bi-alphabetism and bilingualism (of the compound kind), two distinguishable sets of symbols are mapped, in perception, onto a common space; in both cases the mapping of one symbol set was acquired on the basis of the other. By these considerations, bi-alphabetism is a limiting case of bilingualism; and we may conjecture, therefore, that nontrivial asymmetries in processing ought to characterize bilingualism much as they do bi-alphabetism. At all events, further investigation into bi-alphabetism should provide insights into the particular problems of bilingualism and to the general problem of the interrelation of separately used symbol manipulating systems.

REFERENCES

Brand, J. (1971) Classification without identification in visual search. *Quarterly Journal of Experimental Psychology 23*, 178-186.

Broadbent, D. E. (1971) *Decision and Stress*. (London: Academic Press).

Brown, J. (1958) Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology 10*, 12-21.

Clark, H. H. (1973) The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior 12*, 335-359.

Coltheart, M. (1975) Iconic memory: A reply to Professor Holding. *Memory & Cognition 3*, 42-48.

Dick, A. O. (1974) Iconic memory and its relation to perceptual processing and other memory mechanisms. *Perception & Psychophysics 16*, 575-596.

Egeth, H., J. Jonides and S. Wall. (1972) Parallel processing of multi-element displays. *Cognitive Psychology 3*, 674-698.

Forster, K. I. and E. S. Bednall. (1976) Terminating and exhaustive search in lexical access. *Memory & Cognition 4*, 53-61.

Gibson, E. J. (1965) Learning to read. *Science 148*, 1066-1072.

Gibson, E. J. (1969) *Principles of Perceptual Learning and Development*. (New York: Appleton Century Crofts)

Haber, R. N. (1969) Information processing analyses of visual perception: An introduction. In *Information Processing Approaches to Visual Perception*, ed. by R. N. Haber. (New York: Holt, Rhinehart & Winston).

Ingling, N. (1972) Categorization: A mechanism for rapid information processing. *Journal of Experimental Psychology 94*, 239-243.

Jonides, J. and H. A. Gleitman. (1972) A conceptual category effect in visual search: O as letter or as digit. *Perception & Psychophysics 12*, 457-460.

Neisser, U. (1967) *Cognitive Psychology*. (New York: Appleton Century Crofts).

Peterson, L. R. and M. J. Peterson. (1959) Short-term retention of individual verbal items. *Journal of Experimental Psychology 58*, 193-198.

Posner, M. I. (1969) Abstraction and the process of retention. In *Psychology of Learning and Motivation*, vol. 3, ed. by G. H. Bower and J. T. Spence. (New York: Academic Press).

Posner, M. I. (1970) On the relationship between letter names and superordinate categories. *Quarterly Journal of Experimental Psychology 22*, 279-287.

Reynolds, A. G. and P. W. Flagg. (1977) *Cognitive Psychology*. (Cambridge, Mass.: Winthrop Publishers, Inc.).

Sperling, G. (1960) The information available in brief visual presentations. *Psychological Monographs 74* (Whole No. 498).

Sperling, G., J. Budiansky, J. G. Spivak and M. C. Johnson. (1971) Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science 174*, 307-311.

Tomić, T. (1975) Statistical properties of Serbo-Croatian language. Proceedings of the eighteenth conference of ETAN, June.

Tversky, A. (1977) Features of similarity. *Psychological Review 84*, 327-352.

Wickens, D. D. (1970) Encoding categories of words: An empirical approach to meaning. *Psychological Review 77*, 1-15.

Bi-alphabetical Lexical Decision[*]

G. Lukatela[†], M. Savić[†], B. Gligorijević[†], P. Ognjenović[†] and M. T. Turvey[††]

ABSTRACT

The Serbo-Croatian language is written in two alphabets, Roman and Cyrillic. The majority of the total number of alphabet characters are unique to one or the other alphabet. There are, however, a number of shared characters, some of which receive the same reading in the two alphabets, and some of which receive a different reading in the two alphabets. Letter strings were constructed, all of which could be given a phonological interpretation in Roman, but only some of which could be given a phonological interpretation in Cyrillic; some of these letter strings had a lexical entry in Roman, some had a lexical entry in Cyrillic, some had a lexical entry--the same or different--in both alphabets, and some had no lexical entry in either alphabet. In three experiments, subjects reading in the Roman alphabet mode decided as rapidly as possible whether a given letter string was a word. Taken together, the experiments suggest that in the lexical decision task, Serbo-Croatian letter strings (where their structure permits) receive simultaneously two phonologic interpretations. Whether or not this phonologic bivalence impedes lexical decision in the assigned alphabet mode depends on whether or not the letter string has a lexical entry in at least one of the alphabets.

INTRODUCTION

Our concern is with the processes involved in recognizing visually presented words. There is a good deal of evidence to suggest that visual word recognition may be mediated by a phonologic recoding (for example, Meyer, Schvanaveldt and Ruddy, 1974; Rubenstein, Richter and Kay, 1975). At the same time, substantial evidence can be found for the contrary view, namely, that word recognition can proceed independently of phonologic recoding by means of a direct mapping between graphemic analysis and the lexicon (for example, Forster and Chambers, 1973; Kleiman, 1975; Green and Shallice, 1976; Marcel and Patterson, in press). Given these observations, it would seem prudent at

---

[*] To appear in *Language and Speech*, vol. 21, 1978.

[†] University of Belgrade.

[††] Also University of Connecticut, Storrs.

Acknowledgment: This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade, and in part by NICHD Grant HD-01994 to Haskins Laboratories. The authors wish to thank Leonard Katz for his helpful advice on data analysis.

this stage in the development of the theory of word recognition to accept both processes as available to the experienced reader. Presumably, whether one or the other is used, or both are used, depends in a principled fashion on the circumstances. In this light, we may consider Figure 1 as a reasonably representative depiction of the procedures that support word recognition and the relations among them (See Meyer et al., 1974; Marcel and Patterson, in press).

To clarify, the model depicted in Figure 1 assumes two relatively independent routes by which the lexicon can be accessed: one route is a direct route from the graphemic description; in the other route, phonological analysis intercedes between the graphemic description and the lexicon. The model separates the lexicon from the semantic space in the manner of Morton's (1970) logogen model and Quillian's (1969) Teachable-Language Comprehender. The contents of the lexicon--the lexical entries--can be thought of as abstract entities that are activated by or matched to appropriate stimulation from the eyes, the ears and the semantic space. Lexical entries have pointers to their respective locations in the semantic space, and one lexical entry is assumed for each entry in the semantic space; thus, homographs will have as many lexical entries as they have meanings. As intimated above, the relation between the semantic space and the lexicon is not unidirectional. The semantic space relates to the lexicon in the sense of priming semantically related lexical entries. The distinction between the lexicon and the semantic space is drawn primarily in terms of organization: in the lexicon, entries are said to be organized according to frequency of occurrence or usage, whereas in the semantic space the entries are said to be organized according to semantic relations.

Insofar as Figure 1 represents a reasonable account of the processes yielding visual word recognition, the experiments reported here examine the depicted model through the use of the special situation that is provided by the popular use of two alphabets--the Roman and the Cyrillic--in Yugoslavia.

The modern Serbo-Croatian orthography was constructed at the beginning of the 19th century. The properties of the modern alphabet are that each letter stands for a phoneme and the phonemic interpretation of each individual letter is largely invariant and unaffected by preceding and following letters and letter clusters. All letters are pronounced; there are no letters which are made silent by context.

Both the Roman and the Cyrillic alphabets possess the above properties, and in many areas of Yugoslavia both alphabets are used by the local population. This situation is due, in part, to the educational system, which teaches both alphabets in the first and second grade and, in part, to the fact that reading materials come in both alphabets. In Eastern Yugoslavia the children are taught to read and write Cyrillic during their first school year, and Roman during their second; in Western Yugoslavia the children learn first Roman and then Cyrillic.

The Cyrillic and Roman alphabets in Serbo-Croatian do not represent two completely independent sets of letters. Serbo-Croatian letters can be divided into four different groups, which are illustrated in Figure 2. Some letters have the same shape and pronunciation in both alphabets. We will refer to these letters as "common letters." The word for aunt, for example, is written TETKA in Roman and in Cyrillic. However, there are also several letters of the same shape that represent, in the two alphabets, different utterances. We will call them "ambiguous letters." The word deer, for example, is spelled CPHA in Cyrillic. However, if CPHA were read as Roman, the pronunciation would be different and the "word" itself would be meaningless. Similarly, one can combine ambiguous and common letters to write words that have one pronunciation and meaning if read as Cyrillic, and a different pronunciation and a different meaning if read as Roman. Finally, the remaining letters are specific either to the Roman or Cyrillic alphabets. We will refer to these as "the uniquely Roman" or "the uniquely Cyrillic" letters, respectively.

It is evident that the relation between the two alphabets is not the same as the relation between the upper- and lower-case alphabets of, say, English. It is also evident from the preceding that Serbo-Croatian provides a special situation for the study of word perception in particular, and reading in general.

The use of two alphabets in the Serbo-Croatian language invites a modification of Figure 1 along the lines suggested by Figure 3. In particular, two largely separate but partially overlapping alphabet spaces are introduced, where the overlap is constituted by the representations of the common letters. The stage of graphemic description in Figure 1 is understood in Figure 3 as the assigning of representations (structural descriptions) in one or the other (or both) alphabet spaces to the letters in the input letter string. These representations in the alphabet spaces can constrain a search through the lexicon without further mediating steps. In addition, they can map onto their respective phonologic descriptions, in which case the search through the lexicon is phonologically constrained. As in our discussion of Figure 1, it is assumed that both kinds of search can occur together. However, the redesigning of Figure 1 to accommodate two largely separate alphabet spaces brings with it the question of how the four routes to the lexicon--two graphemic and two phonologic--relate in the processing of Serbo-Croatian letter strings.

The experiments reported here are directed at lexical decision. A subject, on presentation of a string of spatially adjacent letters, is required to respond whether the string is a word or not. The minimal form of this procedure can be referred to as the single lexical decision task. A more complex form presents two letter strings, spatially separated, at the same time and requires the subject to respond "yes" if both letter strings are words and "no" otherwise (Meyer and Schvaneveldt, 1971). This procedure might be referred to as the paired lexical decision task; it is used when the relation between letter strings is of interest (see Meyer et al., 1974). Two of the present experiments (Experiments I and III) employ a variant of the paired lexical decision task in which two (related or unrelated) letter strings are presented in succession (rather than simultaneously) and in which the subject must make two successive lexical decisions, one on the first letter string and one on the second. The remaining experiment (Experiment II) uses a single lexical decision task.

Consider lexical decision from the perspective of the Roman mode, that is, from the perspective of whether a string of letters is a word when read in the Roman alphabet. Table 1 identifies eight types of letter string (LS) composed from the Roman alphabet and the correct lexical decision to each string in the Roman mode. A letter string that is constructed from Roman letters is, in the first place, a string in which there are no uniquely Cyrillic letters and, in the second place, a string in which there are letters common to the two alphabets and sometimes letters that are ambiguous (see Figure 2). Table 1 demonstrates that of the letter strings constructed from the Roman alphabet: (1) all can be given a phonological interpretation in Roman (P_R), but only some can be given a phonological interpretation in Cyrillic (P_C); (2) some can have a lexical entry when read as Roman (L_R); some can have a lexical entry when read as Cyrillic (L_C)--even when they do not have a lexical entry when read as Roman--and some can have a lexical entry in both alphabets.

An examination of lexical decision on the letter strings of Table 1 should reveal the relation between accessing the lexicon graphemically and accessing the lexicon phonologically.

EXPERIMENT I

The first experiment explores several relationships in the paired lexical decision task. It seeks to replicate the observation of a priming effect (Meyer and Schvaneveldt, 1971): the lexical decision on a letter string that composes a word is facilitated if the preceding letter string is a semantic relative (Fischler, 1977). Additionally, and more important, the first experiment examines the influence of alphabet ambiguity on lexical decision. Suppose the reader is reading in Roman, that is, accepting and rejecting letter strings as words in Roman, then we can ask whether the latency of decision on any given string will be affected by the fact that the string is a word if read in Cyrillic. To anticipate the design of the experiment: a subject operating in the Roman alphabet mode will be confronted, on some small proportion of the trials, by letter strings that happen to be words in the Cyrillic alphabet mode, but may or may not be words in the Roman alphabet mode.

Method

Subjects. Twenty students from the University of Belgrade Faculty of Philosophy served voluntarily as subjects. All the students had normal or corrected to normal vision, all received their elementary education in Eastern Yugoslavia, and none had had previous experience with visual-processing experiments. One subject was eventually dropped from the analysis because of too many responses in excess of 1500 msec.

Materials and Design. Letraset black uppercase Roman letters (Helvetica Light, 12 point) were used to prepare the letter strings. A string of three to six letters arranged horizontally at the center of a 35 mm slide represented a word or a nonword in the Roman alphabet. The criterion for choice of words was that they belonged to the vocabulary of elementary school children. From published word frequency data for Serbian children (Lukić, 1970), words from the midfrequency range were chosen; too frequent words and too rare words were avoided. In addition, for both word strings and nonword strings, rare consonant clusters were avoided.

The letter strings were grouped into pairs such that either member of a pair could be a word or a nonword. All in all, there were eight different types of pairs, and these are given in Table 2 along with the proportion of trials on which each type appeared in the experiment.

First consider Types 1 and 2. The first and second members of a pair were LS1 and LS1 (see Table 2) for both pair types. In short, those were word/word pairs in the Roman alphabet that were unclassifiable in the Cyrillic alphabet. In Type 1, the two letter strings were associatively related--in Type 2, they were not. Associative norms are not available (to our knowledge) in Serbo-Croatian, so associated and nonassociated pairs were determined by a panel of native Yugoslavians. In contrast with the research of Meyer, Schvaneveldt and Ruddy (1975), different sets of letter strings were used to construct the associated and nonassociated pairs. When a single set of letter strings is used for this purpose, care must be taken in assigning subjects to pairs so that a given subject never sees the same letter string twice. Thus, half the subjects must see half of the Type 1 pairs and the noncorresponding half of the Type 2 pairs; the other half of the subjects then see the other halves of the Type 1 and Type 2 pairs. While this design strategy has the advantage of permitting the comparison of the same letter strings in the associated and nonassociated cases, there are complications in analyzing the data according to the strictures suggested by Clark (1973) (see Meyer et al., 1974; Scarborough, Cortese and Scarborough, 1977).

Type 3 pairs were composed from letter strings of types LS8 and LS1, that is, they were nonword/word pairs in Roman but unclassifiable (unreadable) in Cyrillic. The words in these pairs were different from the second words in the Type 1 and Type 2 pairs. The Type 3 pairs will provide a further but limited control for the Type 1 pairs and the appropriate control for the Type 4 pairs. Type 4 pairs are composed from letter strings of type LS8 and LS3, that is, nonword/word pairs in Roman and unclassifiable/word pairs in Cyrillic. The significant feature of the second letter string of each Type 4 pair is that the Roman reading and the Cyrillic reading specify different words. In short, the second member of Type 4 pairs is a word in both alphabets. For example, CEH means "bill" in Roman and "shadow" in Cyrillic. A comparison of Type 3 and Type 4 pairs permits a determination of whether accepting a string as a word is facilitated by the string's lexical membership in both alphabets.

Type 5 and Type 6 pairs were, respectively, LS8, LS6 and LS1, LS6. That is to say, Type 5 pairs were nonword/nonword in Roman and unclassifiable/word in Cyrillic. An examination of responses to the second members of these pairs will permit the determination of whether rejecting a string as Roman is affected by the fact that the string has a lexical entry in Cyrillic. The controls for Type 5 and Type 6 pairs are provided by Type 7 and Type 8 pairs. Type 7 pairs are nonword/nonword (LS8/LS8) in Roman and unclassifiable in Cyrillic. Type 8 pairs are word/nonword (LS1/LS8) in Roman and unclassifiable in Cyrillic.

Our intention was to have the subject operate in the Roman alphabet mode. We sought to achieve this by creating a context (as opposed to giving an instruction) in which all letter strings were readable as Roman and in which very few letter strings were readable as Cyrillic. There were never any uniquely Cyrillic letters. Strings that were readable in Cyrillic were constructed from the letters common to the two alphabets. A subject saw 72 pairs in the experimental session, that is, 144 letter strings. Of these 144 letter strings, only 27 contained ambiguous characters. These 27 were the only strings that could be read as Cyrillic and they only occurred as second members of a pair.

The 72 pairs seen by a subject were presented in four blocks. In each block the pairs of each type were presented in a pseudo-random order. The sequence of blocks was balanced across subjects according to a Latin square design. The same string of letters was never judged more than once by a subject.

Procedure

The subject was seated at a three-channel tachistoscope (Scientific Prototype, Model GB). The subject was instructed to focus on the fixation point in the center of a preexposure field that was present at all times except during presentation of a letter string. An auditory warning signal preceded the first letter string in a pair. Onset of the letter string triggered an electronic counter that was stopped when the subject pressed either one of two buttons on a response panel in front of him. Both hands were used. Both thumbs were placed on a telegraph key button close to the subject and both forefingers on another telegraph key button two inches further away. The subject depressed the closer button (thumbs) if the letter string was a Roman nonword, and the other further button (forefingers) if the letter string was a Roman word. As soon as a button was depressed, the first letter string of a pair was replaced by the second. When the second letter string was presented, another electronic counter was triggered. The subject now judged whether the new string of letters was a word or a nonword and again made his answer by pressing the telegraph keys in the manner described. Regardless of the subject's response time, the second letter string in each pair was always automatically replaced after 1500 msec by the preexposure field.

Results and Discussion

For all analyses, only the response latencies and errors with respect to the second letter strings were considered. Data were excluded from trials on which the response to the first letter string was incorrect. Incorrect classifications and correct classifications that exceeded 1500 msec were defined as errors. The basic datum was the reaction time (RT) for each subject for each type of stimuli. Table 2 summarizes the results of the experiment.

There are two main aspects of the data. First, the latency of recognizing that the second letter string was a word was significantly affected by the associative relation between the two strings; precisely, where the first string was an associate of the second, lexical decision on the second was enhanced (see Meyer et al., 1975). The mean difference between Type 1 and Type 2 second-string latencies was 92 msec, $F'(2,25) = 10.01$, $p < .001$ (see Clark, 1973). A similar relation clearly holds between Type 1 and Type 3 second string latencies (see Table 2).

Second, it is evident from Table 2 that a letter string that was nonsense in Roman but a sensible Serbo-Croatian word in Cyrillic was rejected as a word with some difficulty. In support of this claim, we may note that rejection latencies for the second letter strings of Type 5 and 6 pairs were generally slower than those for the second letter-strings of Type 7 and 8 pairs. We cannot assess the significance of this contrast because of the enormous error rate that accompanied performance on Types 5 and 6. However, this error rate is instructive. A Wilcoxon signed-ranks test contrasting the proportion of correct second-string responses to Type 5 and 6 pairs with the proportion correct to Types 7 and 8 pairs proves significant ($T_{17} = 2$, $p < .01$). In approximately 20 percent of the trials containing a letter string that was a nonword in Roman but a word in Cyrillic, subjects responded (incorrectly from the perspective of the experiment) that the letter string in question was in fact a word. In approximately 10 percent of the trials containing Roman nonword/Cyrillic word letter strings, correct responses (that is, rejections) took in excess of 1500 msec. In contrast, for the case of letter strings that were nonwords in Roman and unclassifiable in Cyrillic (that is, Type 7 and 8), only five percent of the trials on average were in error in the sense of the string being classified as a word rather than as a nonword. For those Type 7 and 8 strings, approximately less than two percent of these trials were correct classifications in excess of 1500 msec. We may assume, therefore, that on at least one-third of the trials in which subjects viewed Roman nonword/Cyrillic word letter strings, the subjects responded to the Cyrillic interpretation of the strings.

There are two ways to regard the latter observations. In the first place, it can be argued that the conditions of the experiment did not successfully induce a Roman alphabet mode. Against this argument, however, is the fact that of the 144 letter strings seen by a subject during the training and test trials, only 27 of them contained ambiguous characters, that is, only 27 strings suggested a Cyrillic encoding. Significantly, none of these strings contained any uniquely Cyrillic letters. Furthermore, we should remark that other than the aforementioned 27 strings, no other letter strings were even readable as Cyrillic--hence, our classification of these strings as neither words nor nonwords in Cyrillic (see Table 1). The point is that by the design of the experiment, there was very little to encourage the reader to lapse, even occasionally, into the Cyrillic mode of processing.

In the second place, we might regard the comparison of Type 5 and 6 pairs with Type 7 and 8 pairs as indicating that although a reader is in the Roman mode, this does not necessarily prohibit the accessing of the lexicon by Cyrillic script. In the model depicted in Figure 1, two routes to the lexicon are described. Are both routes usable by the Cyrillic version of a letter string when that string is being treated as Roman? Of course, there is nothing in our data that permits an acceptable answer, but let us, for the time being, entertain the following argument: to be in the Roman mode means, essentially, to apply the grapheme-to-phoneme mapping rules that befit the Roman alphabet and its allied orthography. On the face of it, simultaneous application of two different grapheme-to-phoneme rule systems seems unlikely, given the necessity of keeping the ambiguous characters from mutually interfering. In short, the argument is that the Roman relevant rules and the Cyrillic relevant rules cannot operate concurrently, for they are mutually incompatible (see Turvey and Prindle, in press).

Consequently, following this argument, when a reader is in the Roman mode, the phonological route to the lexicon is not open to Cyrillic script. If the Cyrillic version of a letter string does access the lexicon when a reader is in the Roman mode, it can only be by way of the graphemic route.

Consider the string POCA that is not a word in Roman. The graphemic description of this string does have a lexical referent since POCA is a word in Cyrillic; thus a graphemically constrained search of the lexicon will yield a positive answer to the question of lexical membership. On the other hand, the phonological description of this string--given that the reader is in the Roman mode--does not have a lexical referent. In consequence, a phonologically constrained search of the lexicon will yield a negative answer to the question of lexical membership. If it is the case that normal word recognition proceeds, at the very least (see Henderson, 1974), along both graphemically constrained and phonologically constrained lines simultaneously, then we can appreciate that for the Yugoslavian, a letter string like POCA is, in terms of lexical membership, an ambiguous string. We may well suppose that it is this conflict between the graphemically determined answer and the phonologically determined answer that gives rise to the large number of errors in Type 5 and 6 pairs. Insofar as these errors are far fewer than correct decisions, we may further suppose that in cases of conflict the lexical decision is preferentially biased toward the outcome of the phonologically constrained search.

Let us now consider the curious outcome for the second letter strings of Type 4 pairs. Each of these strings is distinguished by the fact that it can be pronounced in both alphabets, though the pronunciations are different, and it is a word in both alphabets, though the words are different. The literature on lexical decision for strings with more than one meaning suggests that strings with multiple meanings are accepted as words faster than strings with a single meaning. The latency difference is pronounced where there is a relatively large difference in number of meanings (Jastrzembski and Stanners, 1975), but marginal where the difference is minimal, such as two meanings versus one (see Clark, 1973; Forster and Bednall, 1976). What makes the present finding curious is that multiple meaning hinders lexical decision and thus runs counter to the more common observation. Positive decisions were over 200 msec slower than those for letter strings that were words only in the Roman alphabet (second strings of Type 3 pairs can be used for comparison), and approximately 23 percent more of the responses were in error. A Wilcoxon signed-ranks test on proportions of correct responses for Type 4 and Type 3 second strings is significant ($T_{15} = 1$, $p < .01$). In short, when a string of letters was a word in both alphabets, responses were very slow (the slowest for all types, see Table 2) and on a relatively large number of occasions, subjects actually decided that these strings were in fact Roman nonwords.

In light of the research on lexical decision and multiple meaning, it would seem that the response tardiness and error cannot be due to the fact that a Type 4 string was a word in both Roman and Cyrillic, but rather to the fact that a Type 4 string could be phonologically interpreted in both alphabets. This interpretation argues against our earlier definition of "being in the Roman mode" as the abrogating of the phonological route to the lexicon by the Roman grapheme-to-phoneme rules. In short, the Cyrillic version of a letter string that is being responded to explicitly as Roman might well access the lexicon by the phonological route.

EXPERIMENT II

The second experiment seeks to determine whether the impaired lexical decision on the second letter strings of Type 4 pairs in Experiment I was due to two lexical entries or to two alternative phonological interpretations. The present experiment focuses on letter strings LS1, LS2 and LS3 (see Table 1). LS1 can be read as Roman but not as Cyrillic and is a word in Roman; LS2 can be read as Roman but not as Cyrillic and is two words in Roman, that is, it is synonymous with a homograph in English; LS3 can be read as Roman and as Cyrillic and it is a word in Roman and a word in Cyrillic. Therefore, while LS2 and LS3 are alike in that they both have two lexical entries, they are dissimilar in that LS2 has but one phonological interpretation, whereas LS3 is phonologically bivalent.

We are reminded that research on English words reveals that lexical decision on homographs is either equivalent to or faster than lexical decision on letter strings with a single lexical entry. Given this fact, we would expect the relation among decision times for the letter strings of the present experiment to be roughly $LS1 \geq LS2 = LS3$. If, on the contrary, two lexical entries impede decision time over one lexical entry--a possible interpretation of the Type 4 results of Experiment I--then the expected relation should be $LS1 < LS2 = LS3$. However, if it is the case that while two lexical entries do indeed facilitate decision time over one lexical entry, this formulation is overridden by the impeding influence of two phonological interpretations, then the relation should be $LS1 \geq LS2 < LS3$.

Method

Subjects. Twenty-two students from the Psychology Department of the University of Belgrade participated as subjects. The majority came from Eastern Yugoslavia.

Materials. Letter strings of three to six letters were composed from Letraset, black uppercase Roman letters (Helvetica Light, 12 point). These were arranged horizontally at the center of 35 mm slides.

Sixty of the letter strings were words: 20 LS1, 20 LS2 and 20 LS3. The sixty nonwords were of the kind LS7 (see Table 1). Each class of words consisted of three subclasses: ten nouns, eight verbs and two adjectives. It is important to note that LS3 is a mix of common and ambiguous letters (see Figure 2). No uniquely Cyrillic letters were used and only the 20 letter strings of type LS3 could be read in Cyrillic; as before, the other strings were unreadable in the Cyrillic mode.

Design and Procedure

Each subject saw the full complement of words and nonwords. Four randomizations of the 120 letter strings were partially counterbalanced across the subjects. Each letter string was exposed for 1500 msec in one channel of the three-channel tachistoscope used in Experiment I. Exposure luminance was $10.3\ cd/m^2$. A timer was initiated at the onset of a slide and was terminated when the subject depressed either the "Yes" buttons or the "No" buttons as described in Experiment I. The first twelve trials were taken as practice trials.

Prior to the experiment each subject was instructed as follows: "Subsequent to the warning signal a string of Roman letters will be presented. Your task is to respond as quickly as possible whether the string of Roman letters is a word or nonsense."

Results

Incorrect responses or responses that were either too fast (less than 300 msec) or too slow (more than 1100 msec) were excluded. For LS1 and LS2 the error rate was approximately 4 percent. For LS3 the error rate was 19 percent. The basic datum was the mean RT for each subject for each type of letter string. The latencies for LS1, LS2 and LS3 were, respectively: 585 ± 53 msec, 564 ± 58 msec and 639 ± 36 msec.

Because of the high error rate associated with LS3, an analysis of latencies is imprudent. Nevertheless, an analysis was conducted, and as suspected, it revealed a significant difference between LS3 and LS2 ($F' = 7.93$, $df = 1, 28$, $p < .01$) and a significant difference between LS3 and LS1 ($F' = 4.4$, $df = 1, 30$, $p < .05$). LS1 and LS2 were not different. A more appropriate test, a Wilcoxon signed-ranks test on proportions of correct responses, yielded a significant difference between LS3 and LS2 ($T_{18} = 5.5$, $p < .01$) and a significant difference between LS3 and LS2 ($T_{19} = 5.5$, $p < .01$). The difference between LS1 and LS2 was not significant.

Discussion

The relation among the three types of letter strings is the same whether we consider latencies or errors: LS1 LS2 LS3. The inference we wish to draw is that decision time to LS3 is impeded, not because it has two lexical entries, but because it has two phonological interpretations. The acceptance of this inference, however, depends on whether we can be convinced that the distinction between LS2 and LS3 is solely the phonological bivalence of the latter.

The letter string of type LS2 has two lexical entries, both of which are accessed through the Roman alphabet; LS3 has two lexical entries, one of which is accessed through the Roman alphabet and one of which is accessed through the Cyrillic alphabet. This distinction between LS2 and LS3 might be important if the lexicon is sensitive to the alphabet by which a lexical entry is accessed. Consider a subject faced in the Roman mode by a string of type LS6. Here he must reject the string as a word, even though it is a word in Cyrillic. Is it that he is able to do so, in part, because the positive, graphemically constrained search is registered as being of Cyrillic origin? That is, there is a tag on the output from the lexicon that indicates the alphabet through which the entry was found. If, in the Roman mode, a graphemically constrained search is successful, but is tagged "Cyrillic," then it can be rejected. The idea that a lexical entry might be tagged according to the alphabet of the string that matched it is reminiscent of the claim in bilingual research that remembered words can be identified as to the language

TABLE 1: Types of letter strings that can be composed from the Roman alphabet.

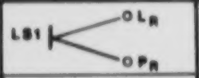
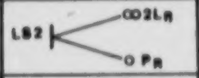
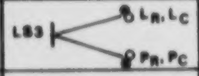
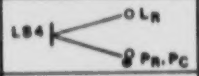
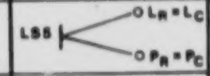
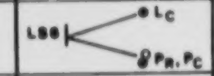
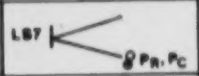
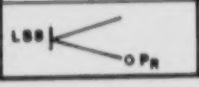
Type of letter string (LS)	Lexical entry (L)		Phonological representation (P)		Symbolic representation	Is it a word in Roman?
	In Roman (L_R)?	In Cyrillic (L_C)?	In Roman (P_R)?	In Cyrillic (P_C)?		
LS1	Yes	No	Yes	No		Yes
LS2	Yes, two	No	Yes	No		Yes
LS3	Yes	Yes	Yes	Yes		Yes
LS4	Yes	No	Yes	Yes		Yes
LS5	Yes	Yes	Yes	Yes		Yes
LS6	No	Yes	Yes	Yes		No
LS7	No	No	Yes	Yes		No
LS8	No	No	Yes	No		No

TABLE 2: Mean reaction times (RTs) of correct responses and proportion of errors for each pair type in Experiment I.

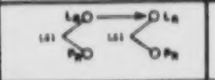
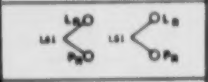
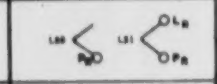
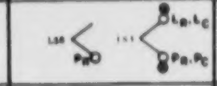
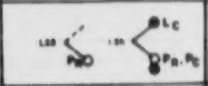
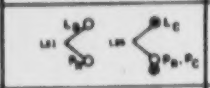
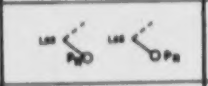
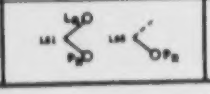
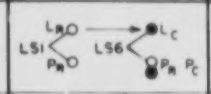
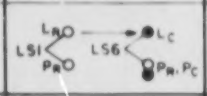
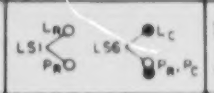
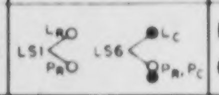
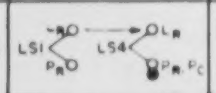
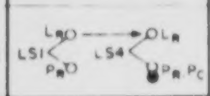
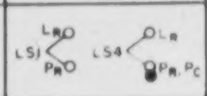
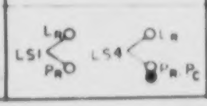
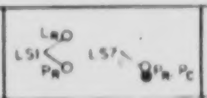
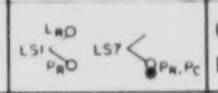
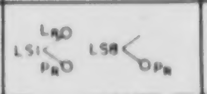
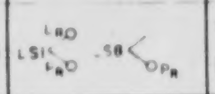
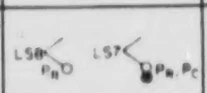
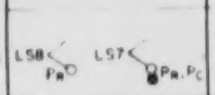
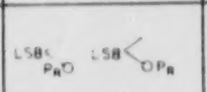
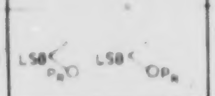
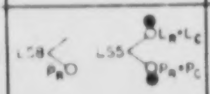
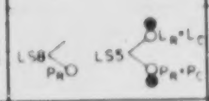
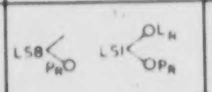
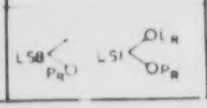
Type of pair	Letter Strings First — Second	Example First — Second	Correct response	Relative frequency	Reaction time (msec)	Percent of errors		
						Regular	Slow	Total
1		OBLAK — KIŠA STENA — KAMEN	Yes	0.11	636 ±67	1.5	1.5	3
2		MAKIT — MLEKO TRAVA — KUĆA	Yes	0.11	726 ±67	2	2	4
3		ŠUPALJ — TOČAK EČANJ — GUMA	Yes	0.20	722 ±84	2	0.5	2.5
4		NUPER — CEN LASET — KACA	Yes	0.08	960 ±184	20	6	26
5		DINAK — PEREP VIGA — POCA	No	0.09	806 ±178	19	13	32
6		GUSKA — TABAN KULA — BETAP	No	0.09	915 ±200	20	16	36
7		ŽITEP — VUNER RILAP — GAPULJ	No	0.11	866 ±135	3.5	3.5	7
8		PČELA — MEREZ LEKAR — DEVIŠ	No	0.20	817 ±125	4	-	4

TABLE 3: Mean reaction times (RTs) of correct responses and proportion of errors for each pair type in Experiment III.

TYPE OF PAIR		LETTER STRINGS	EXAMPLE	CORRECT RESPONSE	RELATIVE FREQUENCY	REACTION TIME (MSEC)	PERCENT OF ERRORS		
		FIRST – SECOND	FIRST – SECOND				REGULAR	SLOW	TOTAL
1	A		OLUJA – BETAP KAIŠ – REMEN	NO	0.06	871 ±53	14.	8.	22.
	B		AŻDAJA – HEMAH STATUA – BAJAP	NO	0.06	831 ±63	26.	4.5	30.5
2	A		STATUA – HEMAH AŻDAJA – BAJAP	NO	0.06	815 ±78	10.	8.	18.
	B		KAIŠ – BETAP OLUJA – REMEN	NO	0.06	837 ±121	3.1	4.7	7.8
3	A		FLAŠA – BOCA LUTKA – PAJAC	YES	0.06	633 ±56	2	2.	4
	B		LOV – HAJKA VEK – EPOHA	YES	0.06	722 ±110	3	1.5	4.5
4	A		VEK – HAJKA LOV – EPOHA	YES	0.06	745 ±103	8.	–	8.
	B		LUTKA – BOCA FLAŠA – PAJAC	YES	0.06	795 ±75	10.5	3.	13.5
5	A		GUSKA – BAMEP KULA – HAJEH	NO	0.06	749 ±91	2.	2.	4.
	B		GUSKA – POHO KULA – EBOC	NO	0.06	728 ±41	4.4	–	4.4
6	A		PTICA – RANTA VESLO – DAZAN	NO	0.11	749 ±52	3.	5.	8.
	B		VESLO – RANTA PTICA – DAZAN	NO	0.11	729 ±52	1.5	3.7	5.2
7	A		ŻAKAT – POHO KULA – EBOC	NO	0.06	763 ±60	6.	4	10.
	B		ŻAKAT – BAMEP KULA – HAJEH	NO	0.06	736 ±83	1.6	1.6	3.2
8	A		NAVET – ŠUĊA MANEK – RAGON	NO	0.11	753 ±54	4	2.	6.
	B		MANEK – ŠUĊA NAVET – RAGON	NO	0.11	783 ±71	2.3	–	2.3
9	A		MEKRE – TETKA ŻUDOS – JAMA	YES	0.11	663 ±38	3	–	3
	B		SIDEK – TETKA NALIR – JAMA	YES	0.11	684 ±47	4.5	0.7	5.2
10	A		AŻULE – MAĆKA VUREM – SESTRA	YES	0.11	632 ±39	1.	–	1.
	B		EČANJ – MAČKA GUFEČ – SESTRA	YES	0.11	658 ±16	1.5	–	1.5

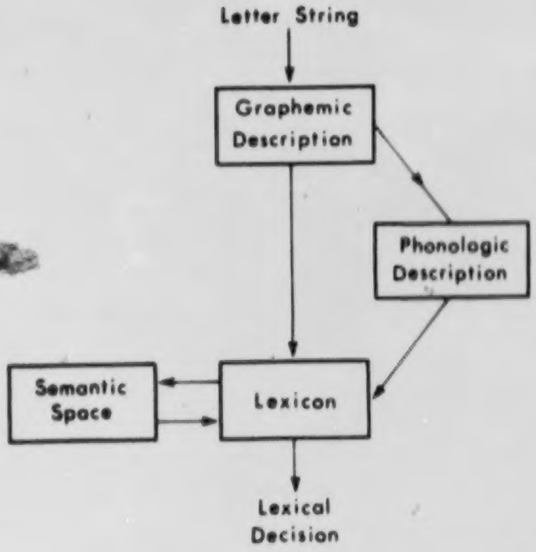

Letter String
Graphemic Description
Phonologic Description
Semantic Space
Lexicon
Lexical Decision

Figure 1: A general model of lexical access.

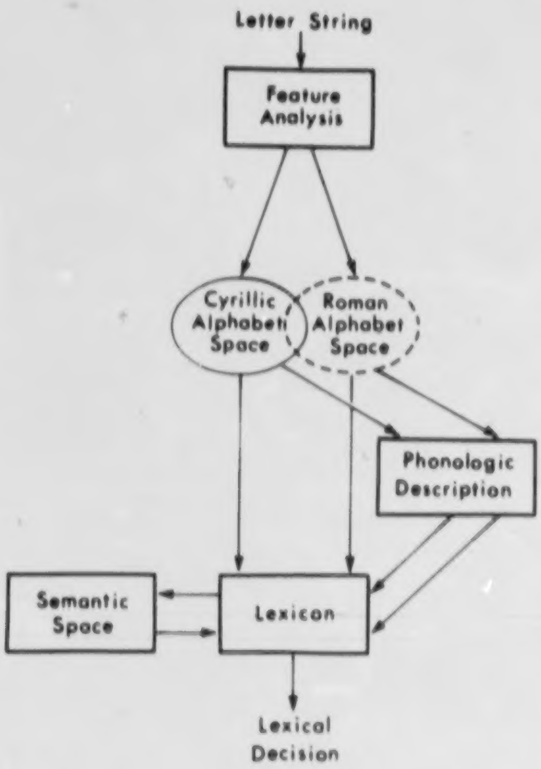

Letter String
Feature Analysis
Cyrillic Alphabet Space
Roman Alphabet Space
Phonologic Description
Semantic Space
Lexicon
Lexical Decision

Figure 3: A modification of the general model of lexical access incorporating the two alphabet spaces.

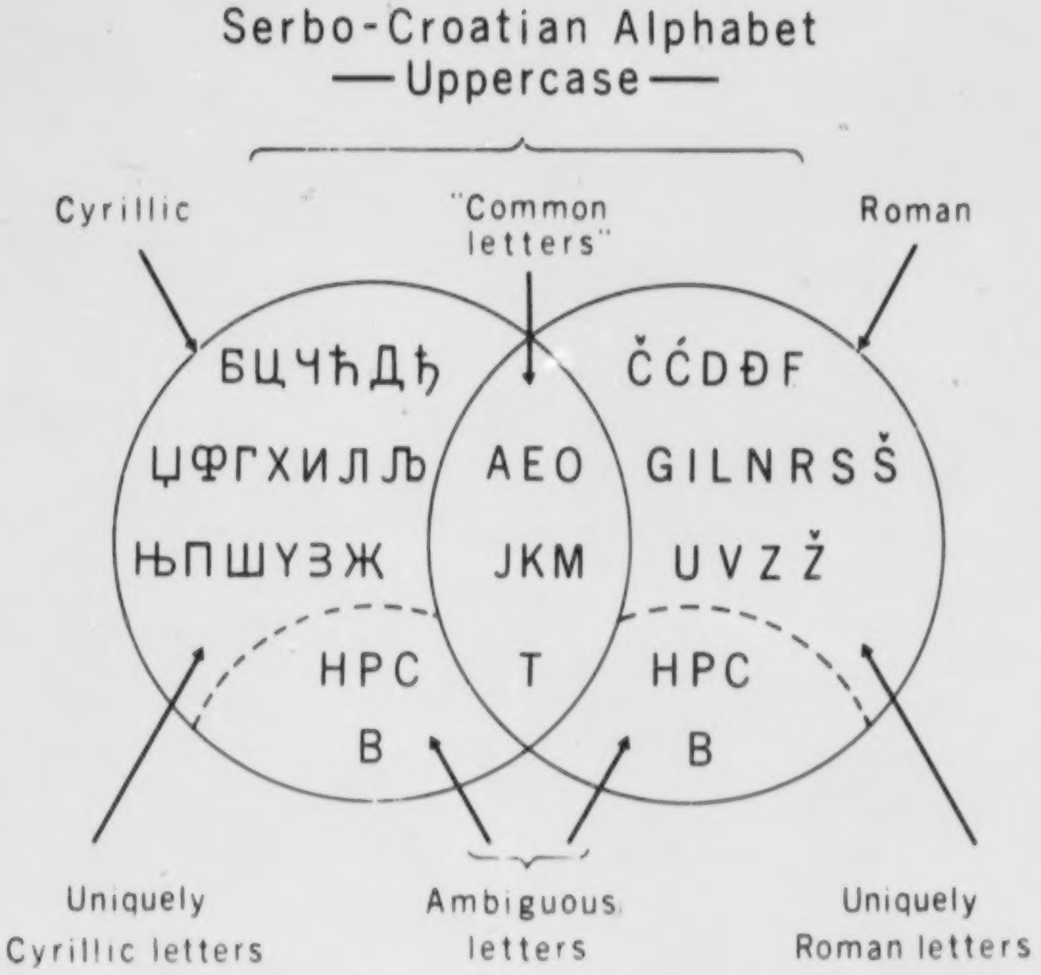

Serbo-Croatian Alphabet
—Uppercase—
Cyrillic
"Common letters"
Roman
БЦЧЋДЂ
ЏФГХИЛЉ
ЊПШУЗЖ
AEO
JKM
T
ČĆDĐF
GILNRSŠ
UVZŽ
HPC
B
HPC
B
Uniquely Cyrillic letters
Ambiguous letters
Uniquely Roman letters

Figure 2: The upper-case letters of the Roman and Cyrillic alphabets.

in which they were received (for example, Saegert, Hamayan and Ahmar, 1975). At all events, we should inquire into a style of processing that distinguishes excited lexical entries by the alphabetic source of their excitation.

Processing the alphabet characters of the Serbo-Croatian language might proceed as follows. Initially, the graphemic features are determined and the resultant feature lists (or structural descriptions) are matched in parallel with the representations of the Cyrillic characters and the Roman characters in the relatively separate Cyrillic and Roman alphabet spaces (see Figure 3). Suppose that matches are found in both alphabet spaces for all characters in the string--as would be true for LS3--then we can imagine that *two* graphemically constrained lexical searches are initiated. In the case of LS3, both of these searches determine a lexical entry; we need only to assume that both of these entries are tagged according to the search that discovered them.

Now we know from the comparison of decision times to LS2 and LS1 that the poor decision performance of LS3 is not due to *two* lexical entries as such. If (for the sake of argument) we rule out phonological bivalence as an influence on decision time, then it must be the case that the poor performance on LS3 is due either to: (1) the fact that there are two different tags, indicating that the lexicon was successfully accessed by both the Cyrillic and the Roman directed search or to (2) the fact that two separately directed searches were conducted simultaneously, or to both (1) and (2).

If conflicts of the kind intimated in (1) and (2) above are the source of the decision time difference between LS3 and LS2 (for LS2 would invite only one lexical search and only one lexical tag, namely the Roman), then they can be investigated with letter strings composed entirely from the common letters (see Figure 2). A letter string so composed (LS5 in Table 1) should, by the preceding reasoning, invite two separately directed lexical searches and yield both a Roman and a Cyrillic tag. A letter string of type LS5, by definition, is common lexically and phonologically to the two alphabets.

The third experiment examines letter strings of type LS5 as part of a general examination of the relationship between lexical entry and phonological bivalence in determining lexical decision time.

EXPERIMENT III

The third experiment is like the first and unlike the second in that it uses the paired lexical decision task. As with Experiment I, the focus is on the decision time to the second letter strings of a pair. For some of the analyses that are of interest in the third experiment, the nature of the first letter strings of a pair is of significance; for most analyses, however, the nature of the first string is irrelevant. In the third experiment, six of the letter strings depicted in Table 1 were examined with LS2 and LS3 excluded. In keeping with the preceding two experiments, the focus of the third experiment is on lexical decision in the Roman mode.

(i) *Priming across alphabets*. It was observed in the first experiment that where the first word of a pair was associated with the second, accepting the second as a word was facilitated. It was also observed that the latency to decide that a letter string was a nonword in the Roman alphabet was retarded if that letter string *was* a word in the Cyrillic alphabet. Suppose that the first string of a pair was a Roman word (and unclassifiable in Cyrillic), and the second string was a Roman nonword but a word in Cyrillic that was *associated with* the (first string) Roman word. Would the latency to reject the second string as a Roman word be further protracted? If priming occurs across alphabets, then we would expect that the first-string's Roman lexical entry would, through the semantic space (see Figure 1), facilitate the second-string's Cyrillic lexical entry and in consequence augment the difficulty in rejecting the second string as a Roman nonword. The relevant comparison is that between Type 1 pairs and Type 2 pairs in Table 3. In both Type 1 and Type 2 pairs, the first strings are LS1 and the second strings are LS6 (see Table 3); but only in Type 1 pairs is there an association between lexical entries.

(ii) *Priming within an alphabet*. A comparison between Type 3 and Type 4 pairs as shown in Table 3 provides a measure of priming within an alphabet. In these pairs the first strings are LS1 and the second strings are LS4; in Type 3 pairs the lexical entries of the successive strings are associated. The comparison between Type 3 and Type 4 pairs differs from the similar comparison of Experiment I, for in the first experiment the second strings were LS1.

(iii) *Significance of phonological ambivalence per se*. If the latency to reject a Roman nonword is impeded by the fact that a letter string can receive an alternative phonological interpretation in Cyrillic, then this impedance should be realized even when the letter string is a nonword in Cyrillic. Experiment I had compared LS6 and LS8 and observed that errors and decision latency on LS6 significantly exceeded these measures on LS8. While LS6 is phonologically bivalent, it also has a lexical entry. The third experiment asks whether a similar relation exists between LS7 and LS8. Neither of these types of letter strings has a lexical entry, but the former (LS7) has two phonological interpretations to the latter's (LS8) one (see Table 1). The relevant comparison is between the second letter-strings of Type 5 and Type 6 pairs and between the second letter-strings of Type 7 and Type 8 pairs (see Table 3).

(iv) *Significance of potential for two lexical searches and two alphabet tags*. The third experiment contrasts the lexical decision on LS5 to that on LS1 in the spirit of the hypotheses developed in the discussion of Experiment II. According to these hypotheses, decision times and errors should relate as [illegible] > LS1. We recall that letter strings of type LS5 are composed entirely from the common letters. Consider then the contrast between LS5 and LS1: LS5 would find a match in both the Roman and Cyrillic alphabet spaces (see Figure 3), but LS1 would find a match only in the Roman space; LS5 would receive a phonological interpretation (the same) whether read in the Roman mode or the Cyrillic mode, but LS1 receives a phonological interpretation only in the Roman mode; LS5 would find a lexical entry (the same) whether read in Roman or Cyrillic, but LS1 has a lexical entry only in the Roman mode. If ambivalence in lexical search or ambivalence in assigning the alphabetic source of lexical outputs is a significant determinant of lexical decision time, then it follows, as argued above, that decision times should relate as LS5 > LS1. The relevant comparison is given by the second letter strings of Type 9 and Type 10 pairs (see Table 3).

Method

Subjects. The participants in the experiment were 40 students from the Department of Psychology at the University of Belgrade. The majority of the students had received their elementary education in Eastern Yugoslavia. They were not unfamiliar with RT experiments.

Materials and Design. Slides containing either a word or a nonword were prepared in the manner described for Experiments I and II. The criteria for choice of words were as described in Experiment I.

There were ten different types of letter string pairs that were of interest; these are shown in Table 3 along with examples of the letter strings and the approximate relative frequency with which each pair type appeared in the trials of the experiment. Other pairs were included to insure a balance between words and nonwords and to keep the proportion of strings readable in Cyrillic at a minimum; these pairs were not analyzed.

First consider pairs of Type 1 and Type 2 whose first and second members are, respectively, letter strings LS1 and LS6. The second members of these pairs, therefore, were nonwords in Roman and words in Cyrillic. In Type 1 pairs, the lexical entry of the second member of the pair was associatively related to the first member of the pair, for example, OLUJA (in Roman) translates as "storm" in English and BETAP (in Cyrillic) translates as "wind" in English. No associative relation holds between members of Type 2 pairs. The pairs of Type 2 were obtained by interchanging first members of the Type 1 pairs. Type 2, therefore, provides a control for the possible priming effect of Type 1.

The first and second members of Type 3 and Type 4 pairs were letter strings of Type LS1 and Type LS4. The second members of these pairs, therefore, were words in Roman and nonwords in Cyrillic. In Type 3 pairs the members were associatively related; for example, FLASA (in Roman) translates as "flask" and BOCA (in Roman) translates as "bottle." No associative relation holds between members of Type 4 pairs; these pairs were obtained by interchanging first members of the Type 3 pairs.

Consider pairs of Type 5 and Type 6. The members of Type 5 pairs were LS1 and LS7 in that order; the members of Type 6 pairs were LS1 and LS8 in that order. Letter strings of Type LS7 can be read in both Roman and Cyrillic, but are nonwords in both alphabets. These letter strings are composed from a mixture of common and ambiguous letters. They were constructed by taking a letter string of Type LS3 and replacing either one or two of the ambiguous consonants in these strings by other ambiguous consonants so as to produce letter strings that were readable and nonsense in both alphabets. Letter strings of Type LS8 are readable only in Roman. They were constructed by taking a letter string of Type LS1 (which is not readable in Cyrillic) and replacing one ambiguous consonant by another to produce a nonsense string.

Other constraints on generating strings of Types LS7 and LS8 should be noted. First, strings should be consonant-vowel sequences as opposed to consonant clusters, in order to increase the likelihood that the ease of giving a phonological interpretation to the strings be equivalent in Roman and Cyrillic. Consonant clusters (for example, CK in CKOJ) differ in ease of pronunciation and frequency of occurrence from one alphabet to the other (thus, CK is easier to say and is more frequent in Cyrillic). Second, care was taken in determining letter strings of Type LS7 so that, on the average, these strings were different by the same number of letters from Roman and Cyrillic words.

Pairs of Type 7 and Type 8 were the same as pairs of Type 5 and Type 6 in all significant respects, except that (1) the first members of a pair were LS8, that is, nonwords in Roman and unclassifiable (nonreadable) in Cyrillic, and (2) the second strings of Type LS8 in Type 8 pairs were different from the second strings of Type LS8 in Type 6 pairs.

Finally, let us consider Type 9 and Type 10 pairs. The first and second members of Type 9 pairs were LS8 and LS5, respectively; and the first and second members of Type 10 pairs were LS8 and LS1, respectively. Only the second members were of interest. Composed solely of common letters, letter strings of Type LS5 were words so chosen as to overlap in frequency of occurrence with the words of Type LS1.

Each of the forty subjects judged 144 letter strings according to the instructions used in Experiment II. Both the instructions and the construction of the letter strings were meant to induce the Roman mode. As before, there were no uniquely Cyrillic letters, and of the 144 letter strings only 32 of them (approximately 23 percent) could be read as Cyrillic.

An individual subject never saw the same letter string twice (see Table 3). A subject received either all the A versions of the ten types of pairs or all the B versions. A subject was assigned either to the A versions or the B versions on order of arrival at the laboratory. The 56 pairs seen by a subject were presented in four blocks. In each block the pairs of each type were presented in a pseudo-random order. The sequence of blocks was balanced across subjects according to a Latin square design.

Procedure. The apparatus, method of response, etc., were identical to those of the first experiment.

Results

The experiment was designed so that for a given pair type, one half of the subjects saw one half of the pairs and the other half of the subjects saw the other half of the pairs. This design guaranteed the general feature that no subject saw the same letter string twice and the particular feature that in the Type 1, Type 2 comparisons and in the Type 3, Type 4 comparisons, the same letter strings could be used for associated and nonassociated pairs. As remarked above, this design imposes difficulties when one is trying to keep the data analysis true to the strictures suggested by Clark (1973); that is, where both subjects and letter strings are treated as "random effects" and reliability of results is computed over both of these sampling domains.

In the kind of analysis[1] we chose, individual quasi-F ratios were computed for comparisons within a comparison. For example, the comparison between Type 3 and Type 4 includes the following sub-comparisons: (a) comparisons in which subjects are the same but letter strings are different: Type 3A versus Type 4A and Type 3B versus Type 4B; and (b) comparisons in which subjects are different but letter strings are the same: Type 3A versus Type 4B and Type 3B versus Type 4A. For some types in Table 3, and for other comparisons we wish to consider, the subcomparisons on different subjects, same letter strings do not exist. In general, then, the subcomparisons will be those where subjects are the same.

---

[1] Katz, L.: personal communication.

The quasi-F ratios for the subcomparisons of a given comparison were considered as random variables whose probabilities have a Chi-square distribution. Suppose that the F' for subcondition X was at the probability level, $p = x$ and the F' for subcondition Y was at the probability level, $p = y$. The new random variables are computed as $r_1 = -2 \ln (x)$ and $r_2 = -2 \ln (y)$ and their sum determined. The Chi-square distribution has 2k degrees of freedom where k is the number of variables (for our example, there are four degrees of freedom). The obtained sum of the new variables is then assessed for significance against the Chi-square value for the corresponding degrees of freedom. The gist of this method is that it asks: Given a set of individual quasi-F ratios with probabilities, p_1, p_2, etc., is it likely that this set of probabilities could have occurred by chance?

Let us consider the results for the comparisons of initial interest, namely, those described in the introduction to the experiment. As with the previous two experiments, the RTs (and sometimes the errors) to the second letter string of a pair were analyzed. First, no F' ratios greater than unity were found for the subcomparisons of Type 1, Type 2 pairs. The high error rate suggests that this negative conclusion be treated with caution. A Wilcoxon signed-rank test on proportions of correct responses was conducted. Of the possible subcomparisons only two were significant: Type 1B versus Type 2B ($T_{13} = 8$, $p < .05$) and Type 1A versus Type 2B ($T_q = 6$, $p < .05$). The error data, therefore, suggest that priming occurred across alphabets.

Second, the subcomparisons of Type 3 and Type 4 pairs revealed the following F' values: for 3A versus 4A, $F'(1,11) = 4.41$, $p < .06$; for 3B versus 4B, $F'(1,18) = 2.45$, $p < .02$; for 3A versus 4B, $F'(1,19) = 7.10$, $p < .02$; for 3B versus 4A, $F' < 1$. These comparisons provide a curious mix, suggesting that priming within an alphabet did and did not occur. In part, these data may reflect the inadequate procedure used for determining associative relation--the use of a small panel of judges rather than associative norms. The availability of the latter for research with English words provides a more reliable basis for selecting pairs of associated words and thus a better opportunity for observing priming.

Third, inspection of Table 3 is sufficient to conclude that there was no difference between the second letter strings of Type 5 and Type 6 pairs (LS7 and LS8, respectively) and no difference between the second letter strings of Type 7 and Type 8 pairs (again LS7 and LS8, respectively). In short, phonological bivalence per se did not seem to retard lexical decision.

Fourth, the comparison between Type 9 and Type 10 was a straightforward F' analysis (the second letter strings of 9A and 9B were identical, as were the second letter strings of 10A and 10B). The analysis proved significant $F'(1,25) = 7.35$, $p < .02$, indicating that latency of response for strings of common letters was slower than the latency for letter strings that did not have the same status in both alphabets.

The lack of difference in lexical decision time to LS7 and LS8 should be contrasted with the significant difference reported in the first experiment for the comparison of LS6 and LS8. The contrast suggests the following hypothesis: phonological bivalence impedes lexical decision only if there is a lexical entry in one or the other alphabet. The confirmation of this hypothesis would lie with showing that, in addition to the already demonstrated equality, LS7 = LS8, the following decision-time inequalities hold: LS4 > LS1, LS6 > LS7 and LS4 > LS5 (see Table 1).

In words, the first inequality is that a letter string that receives a phonological interpretation in each alphabet and has a lexical entry in Roman should be accepted as a Roman word more slowly than a letter string that similarly has a lexical entry in Roman but receives a single phonological interpretation (in Roman). The following subcomparisons of the present experiment provide the appropriate test: 4A with 10A and 4B with 10B. The individual analyses were highly significant, respectively, $F'(1,12) = 8.51$, $p < .01$, and $F'(1,20) = 9.98$, $p < .01$, yielding, by the method described above, $\chi^2(4) = 18.42$, $p < .003$. On the average, decision time to LS4 was 115 msec in excess of decision time to LS1. Clearly, the sought-after relation, LS4 > LS1, holds.

In words, the second relationship (LS6 LS7) is that a letter string that receives a phonological interpretation in each alphabet and a lexical entry in Cyrillic should be rejected as a Roman word more slowly than a letter string that receives two phonological interpretations but has no lexical entry in either alphabet. The following subcomparisons of the present experiment provide the test: 2A versus 5A and 2B versus 5B. The individual analyses were, respectively, $F'(1,16) = 4.22$, $p < .06$ and $F'(1,15) = 7.03$ $p < .02$, yielding $\chi^2(4) = 13.59$, $p < .01$. On the average, decision time to LS6 exceeded that to LS7 by 76.5 msec. The second of the two sought-after relations, LS6 > LS7, would appear to hold. Caution is induced by the relatively high error rates; favoring the conclusion, however, is the fact that the error difference between LS6 and LS7 is in the same direction as the latency difference.

Prior to considering the third desired relationship, namely, LS4 > LS5, let us look analytically at the finding that decision latency to the second letter-strings (LS5) of Type 9 pairs was slower than the decision latency to the second letter-strings (LS1) of Type 10 pairs. In view of the discussion that concluded Experiment II, we should interpret the slower decision time for LS5 as indicative of either a conflict produced by two separately conducted lexical searches or by the assignment of two alphabet tags to the determined lexical entry. While significant, the latency difference between LS5 and LS1 was not that great, a matter of only 28.5 msec. The magnitude of the difference restrains us from concluding that the slower latency to LS5 is evidence against the hypothesis that, with reference to LS3 (that is, letter strings that have two different phonological interpretations and two different lexical entries), the source of impedance in lexical decision is phonological ambivalence rather than a conflict in lexical search or alphabet tagging.

From other research that we have conducted (Lukatela, Savic, Ognjenovic and Turvey, 1978), we have good reason to believe that for Yugoslavian readers indigenous to Eastern Yugoslavia, there is a bias toward regarding common letters as essentially members of the Cyrillic alphabet. The majority of the subjects in the present series of experiments were from Eastern Yugoslavia. This would mean, perhaps, that in the present experiment there was a tendency, however slight, for subjects to regard letter strings of the LS5 type as non-Roman. If so, then a latency difference between LS5 and LS1 might be expected. At all events, we can better appreciate the importance of contrasting LS5 and LS4. The LS4 type is phonologically bivalent but has a single lexical entry in Roman; LS5 is not phonologically bivalent but it similarly has a single lexical entry, one that can be assessed through either alphabet. If lexical decision is slowed primarily by the fact that a lexical entry can be found and/or tagged through both alphabets, then the acceptance latency for LS5 should exceed that to LS4. If, on the other hand, lexical decision is slowed primarily by phonological bivalence contingent upon the presence of a lexical entry in one or the other alphabet, then the acceptance latency to LS4 should be greater than that to LS5. The relevant comparisons are: 4A with 9A and 4B with 9B. Respectively, the analyses revealed that $F'(1,13) = 3.6$, $p < .08$ and $F'(1,16) = 5.2$, $p < .03$, yielding $\chi^2(4) = 12.06$, $p < .02$. The results of the comparison permit the claim that the inequality, LS4 > LS5, holds; the above hypothesis is thereby verified.

This concludes the analysis and discussion of Experiment III, but two points of general concern to this experiment, and the others, deserve comment. First, while the analysis proposed by Clark (1973) has been applied throughout, there are a number of places where its application necessitates a conservative evaluation of the results. The point of Clark's arguments concerning the analysis of experiments using words as stimuli is that the word-sample chosen may not permit a generalization of the results beyond that sample--hence Clark's advocation of treating words as a random effect, rather than as a fixed effect in the analysis. For a number of the analyses reported in the present paper, the words comprising the experimental sample constituted a significant proportion of the total number of words meeting the specified criteria. In short, we could, in a number of places, have treated words as a fixed effect, thereby enhancing the possibility of a significant outcome.

Second, comparisons were sometimes made in the present series of experiments between conditions that differed not only in the variable of interest, but also in whether the correct response to the first and second letter strings in a pair was the same or different. Where the correct response to the successive strings in a pair was the same, a facilitation of response to the second might be expected. However, inspection of Tables 2 and 3 suggests that such facilitation did not occur and therefore could be ruled out as a source of confusion in the present data. With regard to Table 2, response latency to LS1 in Type 2 pairs (Yes-Yes) did not differ from response latency to LS1 in Type 3 pairs (No-Yes); with regard to Table 3, compare pairs of Type 6 (Yes-No) and Type 8 (No-No) and pairs of Type 5 (Yes-No) and Type 7 (No-No); and finally, returning to Table 2, a comparison of pairs of Type 7 (No-No) and

Type 8 (Yes-No) reveals a difference in the direction opposite to a facilitation prediction.

CONCLUDING REMARKS

It has been assumed that by experimental design and by instruction, a subject could be seduced into one of the two possible alphabet modes, specifically the Roman mode, and that the subject remained true to the Roman mode throughout the presentation of the letter strings. It is, of course, a strong possibility that any given subject may have swayed between modes during the course of an experiment and that subjects differed in the degree to which they adhered to the assigned mode. That is, with respect to some letter strings, the attitude of a subject was that of a Roman reader, and with respect to other letter strings, the subject's attitude was that of a Cyrillic reader. If true, we would expect that on some trials a subject's behavior would be consistent with the Cyrillic reading of a letter string rather than the Roman reading. This would contrast with the claim that on any given trial, any given subject assigned both phonological readings simultaneously. Let us see if we can disarm this mode-switching argument.

The lessor to be learned from the error rates to LS1 (see Tables 2 and 3) is that if a subject is switching modes, he or she does not adopt a mode prior to and impervious to a given letter string. It would seem that a letter string's structure must be discerned as able to support the nonassigned alphabet mode for that mode to be realized. The LS1 can be read in Roman but not in Cyrillic. If subjects adopted the Cyrillic mode indifferent to the structure of a letter string (and prior to the string's presentation), then we should expect the error rate on strings of type LS1 to be large and equivalent to that on type LS4; that, most obviously, was not the case. We might wish to argue, therefore, that a typical subject's strategy was as follows: the orthography of a given letter string was discerned as supporting both Roman and Cyrillic readings and then *one* of the two alphabet modes was engaged to give the letter string a phonological interpretation with the chosen mode varying across trials. On this strategy we should expect decision time for LS3 to differ nonappreciably from decision time to LS1 (see Table 1). According to the aforementioned strategy, whatever alphabet mode the subject engages, the lexical quest will be positive and, presumably, as rapid as that for LS1--a case of a single phonological reading and a single lexical entry. The evidence, we are reminded, is to the contrary: LS3 decision time was appreciably slower than LS1 decision time (see Table 2 and Experiment II).

The kind of mode-switching 'model' considered in the preceding remarks is one that assumes mode switching between trials. While there is reason to doubt this kind of mode switching, there remains the possibility of mode switching *within* a trial. Argument must rest with this point, however, for there are, in theory, an indefinite number of plausible within-trial mode-switching models--some of which would yield the pattern of obtained results and some of which would not. In the absence of any (presently discernable) significant constraints on the construction of such models, we consider the enterprise of doing so ill-advised.

We may as well suppose, therefore, that the data of the present series of experiments can be taken at face value, that is, as indexing the influences of the Cyrillic related phonology on "reading" letter strings in the Roman mode. What is to be made of the term "mode" in the present context? As generally used, it is a slippery term (see Turvey and Prindle, 1978). Assume that it refers to the how of processing. (In contrast, "mode" could refer to the what of processing, for example, speech material versus nonspeech material.) Evidently, to be in the Roman mode does not mean, as proposed above, that the phonologically-mediated route to the lexicon is abrogated by the Roman grapheme-to-phoneme rules. That route, apparently, can be shared and, perhaps, without liability. Indeed, the reading we are giving to the present data is that in the lexical decision task, the ascription of phonological interpretation is obligatory and that a letter string--if its structure permits--will receive both the Roman and the Cyrillic phonological interpretations. Without going into detail, the notion of "being in the Roman mode" seems to refer to a selective operation that is late, rather than early, in processing--much like the claim made for selective attention by some students (for example, Norman, 1968) of the phenomenon who locate attention subsequent to a fairly complete pattern recognition process. One possibility is that to be in the Roman mode means that the link between the lexicon and the semantic space, as depicted in Figures 1 and 3, is prohibited for the Cyrillic processing of a letter string. Experiment III provided some evidence counter to this interpretation (the priming across alphabets), but further experimentation is required.

All things considered, we take the bottom line of the present series of experiments to be this: in the lexical decision task, Serbo-Croatian letter strings (where their structure permits) are ascribed, simultaneously, two phonological readings; and whether or not this phonological bivalence impairs lexical decision in the assigned alphabet mode depends on whether or not the letter string has a lexical entry in one of the alphabets. The full implication of this latter result for a general theory of word recognition must await subsequent investigations.

REFERENCES

Clark, H. H. (1973) The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior 12, 335-359.

Fischler, I. (1977) Associative facilitation without expectancy in a lexical decision task. Journal of Experimental Psychology: Human Perception and Performance 3, 18-26.

Forster, K. I. and S. M. Chambers. (1973) Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior 12, 627-635.

Forster, K. I. and E. S. Bednall. (1976) Terminating and exhaustive search in lexical access. Memory & Cognition 4, 53-61.

Green, D. W. and T. Shallice. (1976) Direct visual access in reading for meaning. Memory & Cognition 4, 753-758.

Henderson, L. A. (1974) A word superiority effect without orthographic assistance. Quarterly Journal of Experimental Psychology 26, 301-311.

Jastrzembski, J. and R. Stanners. (1975) Multiple word meanings and lexical search speed. Journal of Verbal Learning and Verbal Behavior 14, 534-537.

Kleiman, G. M. (1975) Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior 14, 323-339.

75

Lukatela, G., M. Savić, P. Ognjenović and M. T. Turvey. (1978) On the relation between processing the Roman and Cyrillic alphabets: A preliminary analysis with bi-alphabetical readers. *Haskins Laboratories Status Report on Speech Research SR-53*, vol. 1.

Lukić, V. (1970) *Active Written Vocabulary of Pupils at the Elementary School Age*. (In Serbo-Croatian). (Belgrade: Zavod za izdavanje udžbenika SR Srbije).

Marcel, A. and K. Patterson. (in press) Word recognition and production: Reciprocity in clinical and normal studies. In *Attention and Performance VII*, ed. by J. Requin. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Meyer, D. E. and R. W. Schvaneveldt. (1971) Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology 90*, 227-234.

Meyer, D. E., R. W. Schvaneveldt and M. G. Ruddy. (1974) Functions of graphemic and phonemic codes in visual word-recognition. *Memory & Cognition 2*, 309-321.

Meyer, D. E., R. W. Schvaneveldt and M. G. Ruddy. (1975) Loci of contextual effects on visual word recognition. In *Attention and Performance V*, ed. by P. Rabbitt and S. Dornic. (New York: Academic Press).

Morton, J. (1970) A functional model for memory. In *Models of Human Memory*, ed. by D. A. Norman. (New York: Academic Press).

Norman, D. A. (1968) Toward a theory of memory and attention. *Psychological Review 75*, 522-536.

Quillian, M. R. (1969) The teachable language comprehender. *Communications of the Association for Computing Macninery 12*, 459-475.

Rubenstein, H., M. L. Richter and E. J. Kay. (1975) Pronounceability and the visual recognition of nonsense words. *Journal of Verbal Learning and Verbal Behavior 14*, 651-657.

Saegert, J., E. Hamayan and H. Ahmar. (1975) Memory for language of input in polyglots. *Journal of Experimental Psychology: Human Learning and Memory 1*, 607-613.

Scarborough, D. L., C. Cortese and H. Scarborough. (1977) Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance 3*, 1-17.

Turvey, M. T. and S. S. Prindle. (1978) Modes of perceiving: Abstracts, comments and notes. In *Modes of Perceiving and Processing Information*, ed. by H. Pick and E. Saltzman. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Lexical Decision for Inflected Nouns*

G. Lukatela†, Z. Mandić†, B. Gligorijević†, A. Kostić†, M. Savić† and M. T. Turvey††

ABSTRACT

Lexical decision times were measured for three grammatical cases of inflected Serbo-Croatian nouns. The grammatical cases occur with different frequencies. Decision times were not related by a unique constant multiplier to the logarithms of the respective case frequencies. The result suggests that a principle of organization in addition to frequency of occurrence is involved in the lexical memory of inflected nouns.

INTRODUCTION

Several investigators have suggested that during reading, the recovery of word information involves a relatively extensive search of lexical memory (for example, Rubenstein, Garfield and Millikan, 1970; Stanners and Forbach, 1973; Forster and Bednall, 1976). Individual words are said to be represented as lexical entries, with the lexical entries ordered by frequency of occurrence. A search of the lexicon, then, might be construed as beginning at the most frequent entry and searching serially through the list of lexical entries, in accordance with the frequency ordering, until the target entry is determined. If there is no entry, then the search is exhaustive (see Forster and Bednall, 1976).

The focus of this paper is the structure of lexical memory for the Serbo-Croatian language in which inflection is the principal grammatical device. Thus for nouns, all grammatical cases in Serbo-Croatian are formed by adding to the root form an inflectional element, namely, a suffix consisting of one syllable of the vowel or vowel-consonant type.

For any given noun the grammatical cases produced by inflection are not equal in their frequency of occurrence. Table 1 is taken from data collected by Dj. Kostić (1965a); it gives the case frequencies for nouns in the singular

---

*To appear in Language and Speech, vol. 21, 1978.

†University of Belgrade.

††Also University of Connecticut, Storrs.

Acknowledgment: This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade, and in part by NICHD Grant HD-01994 to Haskins Laboratories. The authors wish to thank Leonard Katz for his helpful advice on data analysis.

that are more frequent than nouns in the plural (74 percent to 26 percent). We see, in short, that for any given noun of frequency of occurrence (f) in the language, the singular nominative form will appear with a frequency of approximately .25f, the singular genitive form will appear with a frequency of approximately .20f, and so on.

TABLE 1: The case frequencies of the Serbo-Croatian nouns in singular.

Case	Symbol	Frequency (percent)
Nominative	$(CF)_{nom}$	24.55
Genitive	$(CF)_{gen}$	19.90
Dative	$(CF)_{dat}$	1.86
Accusative	$(CF)_{acc}$	13.52
Instrumental	$(CF)_{ins}$	4.70
Locative	$(CF)_{loc}$	8.79

How might the nouns of an inflected language such as Serbo-Croatian be organized in lexical memory? One hypothesis is that each grammatical case for each noun receives a lexical entry and these lexical entries are ordered according to frequency of occurrence. An alternative hypothesis concurs that each grammatical case for each noun receives a lexical entry, but stresses that frequency is not the sole principle of organization. For any given noun the nominative singular is the most frequently occurring grammatical case and it is that which is learned first. The alternative hypothesis might take the form that nominative singulars are ordered in the lexicon according to frequency of occurrence, but that the other grammatical cases for any given noun are subentries to the noun's nominative singular, and these subentries are organized by some principle other than frequency. A simple prediction follows. If the first hypothesis is correct, then the lexical decision ("is this a word?") latencies for the different grammatical cases of a noun should be determined by frequency of occurrence. However, the lexical decision latencies need not be so determined if the second hypothesis holds.

The present experiment examines Serbo-Croatian nouns from the mid-range of word frequencies (Dj. Kostić, 1965b). For each noun, three singular cases were considered: nominative, locative and instrumental. If the noun occurs with frequency (f), then by the first hypothesis, decision time should be related by a unique constant multiplier to the corresponding logarithms of the proportional frequencies, .25f, .09f, .05f (corresponding to the nominative, locative and instrumental, respectively). By the second hypothesis, decision time to the nominative singular should be fastest, but the relation among the decision times should not be accountable for by the proportional frequencies.

METHOD

Subjects

Thirty-nine students from the Psychology Department of the University of Belgrade participated in the experiment. They were experienced with reaction time procedures.

Materials

The nouns were selected according to the following criteria: (1) easy to read aloud; (2) easily imagined (concrete nouns); (3) only one meaning that was invariant for all grammatical cases; (4) written as alternations of single consonants and vowels. One hundred twenty words were selected for the experiment: 57 nouns in masculine, 52 in feminine and 11 neuter, corresponding to the proportion of genders in the Serbo-Croatian language.

Nonwords were generated as follows. The selected 120 words were listed according to frequency of occurrence. Every other three words in the list were converted into nonwords. For nominatives and locatives this was done by changing the first letter. For example, the noun in nominative "KIŠA" (English: rain) was transformed into the nonsense letter string "GIŠA." In the locative this noun is KIŠI; the nonsense form was LIŠI. For instrumentals, half of the nonwords were produced by changing the first letter and half by changing the last letter or the last two letters. This was done to minimize the influence of the idiosyncratic instrumental endings. For purposes of subsequent analysis it should be noted that the dative and locative for all genders have identical codings and are indistinguishable in the absence of sentential context. Similarly, in the singular, nominative and accusative for masculine and neuter gender are of identical form, whereas in the singular of the feminine gender, the nominative and the accusative are different. In Serbo-Croatian, for all genders, the instrumental is the only unequivocal grammatical case in either the singular or the plural.

The words and nonwords were presented as lower case, printed Roman letters (Helvetica Light, 12 point), horizontally arranged at the center of 35 mm slides.

Procedure

Each of the 120 letter strings was exposed for 1500 msec in one channel of a three-channel tachistoscope (Scientific Prototype, Model GM) illuminated at 10.3 cd/m^2. Both hands were used in responding to the stimuli. Both thumbs were placed on a telegraph key button close to the subject and both forefingers on another telegraph key button two inches further away. The closer button was depressed for a "No" response (the string of letters was not a word), and the further button was depressed for a "Yes" response (the string of letters was a word).

Latency was measured from stimulus onset. The total session lasted for half an hour with a short pause after every eighteen slides.

Design

One hundred twenty stimuli were presented to each subject. Twelve stimuli were used for practice, but were not taken into the final analysis. The subjects were divided into three groups in order to exclude the possibility that the same word, though in different grammatical cases, could be presented to the same subject. Hence, a subject saw one-third of the words and nonwords in nominative, one-third in dative, and one-third in instrumental.

Results

The reaction time of each subject to each stimulus was the basic datum for the analysis. If the subject gave an incorrect answer, his average latency for the given class of stimuli replaced the missing data. The number of incorrect decisions was relatively small (2.4 percent); those responses that were either too fast (less than 300 msec) or too slow (more than 1500 msec) were also considered as errors. The data are summarized in Figure 1.

The reaction times for three inflected forms within each word were compared. A given word in a particular grammatical case was seen by a third of the total number of subjects and, therefore, for the purpose of analysis, the words were divided into three groups of eighteen words each.

The analysis of variance included the three factors: fixed factor--grammatical case, random factor--subjects and random factor--words. A group of thirteen subjects was nested under a particular grammatical case, while the same eighteen words appeared in three inflected forms under the respective treatments.

The differences between the nominative on one side and the instrumental and locative on the other are statistically significant (see Clark, 1973) $F'(1,32) = 5.4$, $p < 0.05$ and $F'(1,35) = 4.05$, $p < 0.05$ respectively, whereas the difference between locative and instrumental was not significant.

Discussion

The results of the experiment demonstrate that, in lexical decision, the latency to nouns in the nominative case is shorter than to nouns in the locative and instrumental cases, and that nonwords take longer to classify than words.

As depicted in Figure 1, the two reaction time plots--one for words and the other for nonwords--display two different patterns. Let us first address the less significant issue of why the latencies to the "instrumental" nonwords were longer than those to the "nominative" and "locative" nonwords. The relative difficulty with the "instrumental" nonwords most probably stems from the fact that the "instrumental" nonsense letter strings were, on average, one letter longer than the other nonsense letter strings. We recall that an "instrumental" nonword was produced by changing one letter in a noun that was grammatically inflected in the instrumental case. We recall, also, that the characteristic ending of the instrumental case in Serbo-Croatian consists of two letters (of the vowel-consonant type) and that the characteristic endings

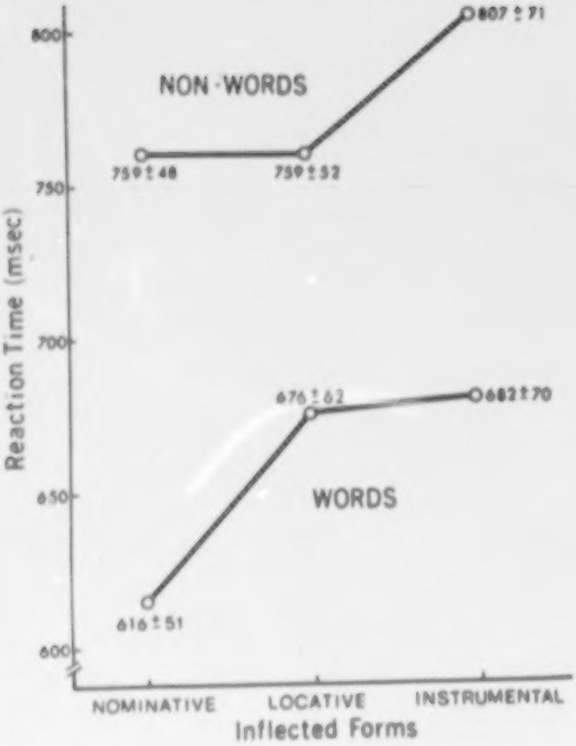


Figure 1: Lexical decision latency as a function of grammatical case.

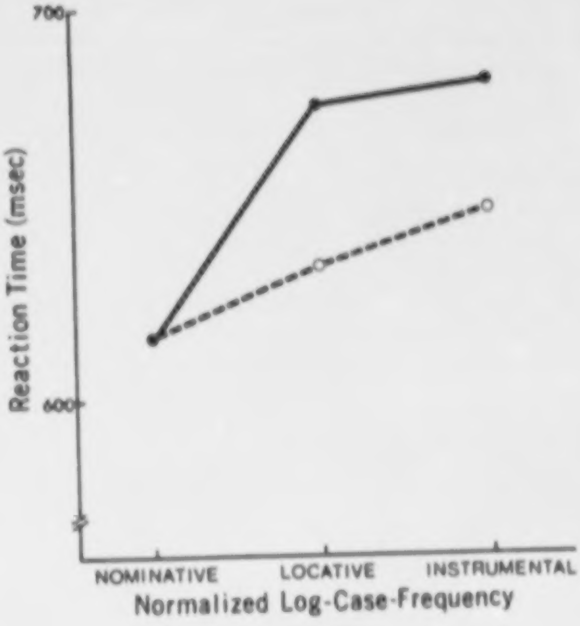


Figure 2: Contrast between obtained (solid line) and predicted (broken line) latencies as a function of normalized log case frequency.

of other cases in singular consist of a single letter (a vowel). As a result of the transformation rules, all nonwords in the experiment were orthographically legal. The "nominative" nonwords were mono- or bisyllables. The "locative" as well as the "instrumental" nonwords were bi- or trisyllables, but each "instrumental" nonword had one letter more than its "dative" mate. These facts and the data in Figure 1 suggest that the effect of number of syllables on lexical decision time for nonwords was not significant. On the other hand, the effect of number of letters in a nonsense string proved to be significant. This finding is in agreement with the results of Forster and Chambers (1973) and Fredericksen and Kroll (1976).

Further comment on the nonword data is unnecessary. Let us focus on the main issue of why the latencies to the inflected words did not follow the general pattern that was predicted by the word frequency effect. In the lexical decision task, the reaction time (RT) is inversely proportional to the word frequency, (f). In the first approximation there is a linear regression of a general form:

$$RT = -A \ln f + B \quad (1)$$

where A and B are the regression coefficients that depend on the number of letters in the word. For English five-letter words, given their frequency of occurrence (Kučera and Francis, 1967), it has been found[1] experimentally that the appropriate numerical values of the regression coefficients are: $A = 17.78$ and $B = 644$.

In the present experiment the average number of letters (when averaged across all nouns in all inflected forms) was about five per word. Therefore, if the reaction time for inflected forms were governed uniquely by the case frequency CF, then the slope of the function relating RT to log CF should be about 17.78, as shown by the dashed line in Figure 2. The zero-intercept of the dashed line, in agreement with our data, was set at $B = 616$ msec.

The experimental data in Figure 2 are represented by black dots and, for convenience, are connected by solid lines. The suggestion is that the solid curve differs from the dashed-line curve not only quantitatively, but also qualitatively. Hence, the plots in Figure 2 suggest that the word frequency effect cannot explain the experimental results.

There is, of course, some theoretical possibility that the numerical value of the slope coefficient A, as plotted in Figure 2, is not appropriate for Serbo-Croatian words. What we need, therefore, is a stronger proof that the experimental data and the data predicted uniquely by the word frequency effect are significantly different for any arbitrary value of the regression coefficients.

---

[1] Katz, L.: personal communication.

The data of Table 1 show that the case frequencies of the nouns in nominative, locative and instrumental relate as follows:

$$(CF)_{nom} > (CF)_{loc} > (CF)_{ins} \tag{2}$$

In a lexical decision task the case frequencies of nominative and accusative for masculine and neuter gender have to be compounded. In the experiment the number of nouns in masculine and neuter gender was sixty-eight, as compared with fifty-two nouns in feminine gender. The joint frequency of occurrence of the unequivocal and equivocal nominative forms, when averaged across all of the one hundred twenty nouns, results in the compounded nominative-accusative case frequency: $(CF)_1 = 31.31$ percent. Similarly, we have also to compound the case frequencies of the locative and dative for all nouns. The compounded locative-dative case frequency is: $(CF)_2 = 10.65$ percent.

If it were true that the mean reaction time and the case frequency were related by equation (1), then between the reaction time to the compounded nominative case $(\overline{RT})_1$ and the reaction time to the compounded locative case $(\overline{RT})_2$, the following hypothetical relation should hold:

$$(\overline{RT})_2 - (\overline{RT})_1 = A \ln \frac{(CF)_1}{(CF)_2} \tag{3}$$

where A is an arbitrary constant; $(CF)_1$ is the compounded case frequency for nominative and accusative, and $(CF)_2$ is the compounded case frequency for locative and dative.

Similarly, for the difference between the mean reaction time to the instrumental, $(\overline{RT})_{ins}$, and the mean reaction time to the compounded nominative $(\overline{RT})_1$, the predicted hypothetical relation would be:

$$(\overline{RT})_{ins} - (\overline{RT})_1 = A \ln \frac{(CF)_1}{(CF)_{ins}} \tag{4}$$

By dividing equation (3) by equation (4) we obtain:

$$\frac{(\overline{RT})_2 - (\overline{RT})_1}{(\overline{RT})_{ins} - (\overline{RT})_1} = \frac{\ln \frac{(CF)_1}{(CF)_2}}{\ln \frac{(CF)_1}{(CF)_{ins}}} \tag{5}$$

If we substitute the numerical values of the mean RTs from Figure 1 into the left side of equation (5), we find that the ratio of the normalized $(\overline{RT})$ difference is:

$$\frac{(\overline{RT})_2 - (\overline{RT})_1}{(\overline{RT})_{ins} - (\overline{RT})_1} = \frac{676-616}{682-616} = 0.91$$

On the other hand, if we substitute the numerical values of the compounded case frequencies as well as the instrumental case frequency into the right side of equation (5) we find that:

$$\frac{\ln \frac{(CF)_1}{(CF)_2}}{\ln \frac{(CF)_1}{(CF)_{ins}}} = 0.56$$

Thus, we conclude that the hypothetical equation (5) is not correct: the left side is numerically about two times as large as the right side.

The preceding mathematical analysis supports the hypothesis that the longer latency to inflected words cannot be accounted for by the difference in the frequency of occurrence of the grammatical forms. We are led, therefore, to the tentative conclusion that frequency of occurrence is not sufficent to capture the lexical organization of the grammatical cases of inflected nouns.

REFERENCES

Clark, H. H. (1973) The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior 12, 335-359.

Forster, K. I. and S. M. Chambers. (1973) Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior 12, 627-635.

Forster, K. I. and E. S. Bednall. (1976) Terminating and exhaustive search in lexical access. Memory & Cognition 4, 53-61.

Frederiksen, J. R. and J. F. Kroll. (1976) Spelling and sound: Approaches to the internal lexicon. Journal of Experimental Psychology: Human Perception and Performance 2, 361-379.

Kostić, Dj. (1965a) Frequency of occurrence of words in Serbo-Croatian, vol. 1, Daily Press. Unpublished. (Belgrade: Institute of Experimental Phonetics and Speech Pathology).

Kostić, Dj. (1965b) The structure of usage value of grammatical forms in Serbo-Croatian: Nouns. (Belgrade: Institute of Experimental Phonetics and Speech Pathology).

Kučera, H. and W. N. Francis. (1967) Computational Analysis of Present-day American English. (Providence, R.I.: Brown University Press).

Rubenstein, H., L. Garfield and J. A. Millikan. (1970) Homographic entries in the internal lexicon. Journal of Verbal Learning and Verbal Behavior 9, 487-494.

Stanners, R. F. and G. B. Forbach. (1973) Analysis of letter strings in word recognition. Journal of Experimental Psychology 98, 31-35.

The Phonetic Plausibility of the Segmentation of Tones in Thai Phonology*

Arthur S. Abramson†

ABSTRACT

In such Southeast Asian tonal languages as Central Thai, the domain of a tone is ordinarily taken to be the syllable, but some linguists have claimed that a segmental representation of the tones best fits the grammar. Thus, the five-way tonal contrast present in the Thai lexicon would be handled by various arrangements of three level tones, underlying which are two binary features. The question is raised as to what kind of phonetic evidence, either in the form of fundamental-frequency contours or perceptual data, would support such a claim. The resulting criteria applied to productions of isolated Thai words and words embedded in sentences fail to provide any direct support for a segmental representation of the tones. In addition, listening tests with controlled variants of fundamental-frequency contours made with a speech synthesizer also fall short of the goal. It is concluded that the phonological arguments for segmentation are weak, that the phonetic data render it implausible, and that the concept is psychologically unconvincing.

INTRODUCTION

The specification of each morpheme in a tone language includes not only a sequence of consonantal and vocalic features, but also a distinctive pitch pattern that is manifested physically in the fundamental frequency of the voice. Linguists have generally analyzed Central Thai (Siamese) as having a five-way tonal contrast, with the syllable as the domain of the tone. There are said to be three level or static tones--mid, low and high--as well as two gliding or dynamic tones--rising and falling.

Some phonologists (for example, Trager, 1957; Leben, 1973; Gandour, 1975) have argued that the holistic treatment of tones in Thai is inherently wrong and should be replaced by a segmental treatment with various sequences of single vowels, double vowels, and final sonorants as the proper domain. While such arguments on the part of Trager (1957) may be a matter of personal taste

---

*To appear in the Proceedings of the XIIth International Congress of Linguists, (Vienna), 29 August-2 September, 1977.

†Also at University of Connecticut, Storrs.

Acknowledgment: This work was supported by NICHD Grant HD-01994 and BRSG RR-5596.

in the manipulation of symbols for the writing of an efficient grammar, the generative treatments must be taken more seriously, since claims are made in this school of thought that the grammar should reflect the speaker's internalized knowledge of his language. By this reasoning, we must suppose that the speaker of Thai stores a lexical item with a dynamic tone as a properly ordered sequence of high and low tones or tonal features.

Linguists with the holistic view of Thai tones have never felt obliged to defend their position. They knew the language well, and it seemed intuitively correct not to segment the tones. This feeling was supported by the native Thai grammatical tradition reflected in the orthography that provides for the correct reading of the tones. Although there is scant literature on children's acquisition of Thai, my own observations and those of others suggest that children learn their basic vocabulary with a tonal contour as an integral part of each item. In fact, children may learn the dynamic tones before the static ones (Sarawit, 1976).

SEGMENTATION OF TONES

The segmentalists argue that consonantal constraints upon the freedom of occurrence of the tones indicate a mapping of each tone onto a segmental base at the level of the underlying form. All five tones may occur contrastively only on syllables that end in a long vowel, or a short or long vowel followed by a sonorant. Except for a few loan words and onomatopoeic terms, a syllable with a short vowel followed by a final stop may take only the high or low tone, while a long vowel followed by a final stop may take only the low or falling tone. In addition, the lexicon includes practically no high or rising tones after certain initial consonants. It is also claimed that tone alterations in compound words are stated in a better formalism with a segmental approach. The general argument rests on the controversial premise that long vowels are sequences of two short vowels.

My thesis here is that a segmental analysis of the tone of Thai is unreasonable and unrealistic. I am not, however, arguing that such an analysis is not appropriate to any language. The most convincing case is one in which all contour tones are obviously derived from underlying sequences, as when vowels undergo sandhi across a morpheme boundary, bringing about a merger of the final static tone of the first morpheme and the beginning static tone of the second morpheme to yield a contour.

Some African languages are said to have a rule of tone copying (Leben, 1973). An inherently toneless syllable takes on the immediately preceding tone. Thus, a toneless element will become high after a high tone and low after a low tone. If, however, the preceding syllable bears a contour tone, the toneless element copies only the final "tone" of the alleged sequence in the contour. The tone-copying rule taken alone as an argument for segmentation succumbs to a natural explanation, which is simply that the pitch movement of the preceding syllable persists in its course through any following element that does not carry a distinctive tone of its own. Even if the latter arguments are accepted, the sandhi feature could lead to a segmental analysis of the tones of those languages anyway, although among these African languages there seem to be some that can be shown to have underlying contour tones (Elimelech, 1974).

If, as it seems, the speaker of Thai learns every morpheme with its tone contour, why must a grammar include complicated rules to express the few consonantal limitations on freedom of occurrence of the tones? These facts are simple and may be seen as part of the speaker's knowledge without letting them force us into an improbable view of lexical entries. In fact, this knowledge has not kept Thai from breaking these "rules" in the tonal treatment of loan words. As for tone alternations and neutralizations in compound words, Gandour (1974) has shown instrumentally that the kinds of examples given by Leben (1973) are by and large untenable.

PHONETIC EVIDENCE

If we believe that the phonology of a language should lead very directly to correct phonetic outputs and auditory percepts, what phonetic evidence would help settle the argument? Would a phonologically disinterested phonetics point to a segmental organization of the tones? A good basis would be acoustic data showing that each of the static tones normally appeared as a level with, perhaps, slight contextually induced perturbations. If each dynamic tone normally appeared as a sequence of these levels with a rapid glide between them, the phonetic evidence would be even more consistent with a segmental analysis. Instrumental investigation of the physiological mechanisms underlying the tones might show segmentation in laryngeal maneuvers or aerodynamic forces. Perceptual evidence might be that static tones are more acceptable when produced as absolute levels rather than movements of fundamental frequency. Also, dynamic tones produced segmentally ought to be more acceptable than mere glides without end-point levels. One more phonetic question is the plausibility of the segmentation of long vowels into two short vowels onto which the tonal segments are mapped. There should be evidence of rearticulation halfway through a long vowel.

Fundamental-frequency contours of Thai tones (Abramson, 1962, 1975; Erickson, 1974) give no acoustic support to the segmental analysis. Although a criterion of relative movement seems to justify the dichotomy between static and dynamic tones (Abramson, 1976), it is nevertheless true that all five tones show much movement. There are no true levels, and the dynamic tones are specified by their direction of movement and not by their end points.

Among the static tones, the fundamental frequency pattern that comes closest to being a true level is that of the mid tone, but even so, it moves upward or downward at both ends or throughout its extent through tonal coarticulation. The low tone starts near the beginning of the mid tone, drops quickly at first, and then falls slowly to the bottom of the voice range. Its early fall distinguishes the low tone from the mid tone. The high tone starts just above the middle of the voice range and, often after a dip, slowly rises. The dynamic tones are exaggerations of the static tones. The falling tone starts just above the middle of the voice range, rises, and then falls abruptly to the middle or bottom of the range. It may thus be better named the high falling tone as contrasted with the low tone, which is a low falling tone. The rising tone starts near the beginning of the mid tone, drops quickly to the bottom of the voice range, then moves abruptly upward. The rising tone is thus really a low rising tone, while the high tone is a high rising tone.

The patterns of laryngeal-muscle activity underlying the contours of the tones of Thai might seem to support a segmental analysis. Such has been Erickson's interpretation of the data in her important dissertation (1976). Using electromyography, she found the activity patterns of a number of laryngeal muscles during the production of the five tones. Two muscles best represent her data. One of them, the cricothyroid, is the principal agent in the control of fundamental frequency. Its contraction stretches and stiffens the vocal folds causing the frequency to rise; when it relaxes, the frequency falls. The other is the thyrohyoid, one of the strap muscles, whose role in the control of fundamental frequency is moot. They contract in association with sharp falls in frequency, but no causal relationship has been demonstrated.

Erickson finds distinctive muscle patterns for the five tones. It is in the dynamic tones that she most readily finds support for segmentation. The rising tone shows a thyrohyoid peak for its initial drop, followed by a cricothyroid peak for its sharp rise, while the falling tone shows a cricothyroid peak first, for its initial rise, followed by a thyrohyoid peak for its sharp fall. The static tones, even when occurring on long vowels, are not obviously to be divided temporally into segments of contraction and relaxation nor, for that matter, do they show uniform patterns throughout, as might be expected in true geminate tones. If one reads support of a segmental view into the complicated muscle data, one is then obliged to reconsider the phonetic integrity of a number of conventionally accepted vocalic and consonantal segments with their temporally resolvable peaks of muscle activity, as in aspirated stops and semi-vowels.

As for perception, some observers hear the static tones as levels, and it is possible that in some instances of these tones auditory averaging of small movements will indeed give the impression of levels; however, it is easy to hear pitch changes most of the time. Indeed, many foreigners have trouble distinguishing between the mid and low tones on the one hand and the mid and falling tones on the other. That is, although experiments in speech perception (Abramson, 1976) do support a dichotomy between tones with large pitch shifts and those without, the term static for the latter is an exaggeration. Although other experiments show that fundamental-frequency levels can be heard as the three static tones by Thai subjects, their acceptability is enhanced when they are synthesized as glides (Abramson, 1975). One can synthesize very acceptable dynamic tones by using continuously changing contours (Abramson, 1962, 1975, 1976), but preliminary work suggests that rapid movements between low and high levels will not yield equally acceptable dynamic tones.

Acoustic data do not enable us to show that the long vowels of Thai are segmentable into sequences of two occurrences of the same vowel (Abramson, 1962, 1974), nor do I know of any electromyographic evidence of rearticulation in long vowels.

CONCLUSION

The arguments for segmentation based on interactions between tones and consonants are too devious and weak to be convincing, and when we turn to phonetic data, the argument becomes even less plausible. I conclude that the traditionally espoused unitary status of the tones of Thai is unshaken.

REFERENCES

Abramson, A. S. (1962) *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments.* (Bloomington: Indiana University, Research Center in Anthropology, Folklore, and Linguistics, Pub. 20).

Abramson, A. S. (1974) Experimental phonetics in phonology: Vowel duration in Thai. *Pasaa 4*, 71-90.

Abramson, A. S. (1975) The coarticulation of tones: An acoustic study of Thai. *Haskins Laboratories Status Report on Speech Research SR-44*, 119-125.

Abramson, A. S. (1976) Static and dynamic acoustic cues in distinctive tones. *Haskins Laboratories Status Report on Speech Research SR-47*, 121-128.

Elimelech, B. (1974) On the reality of underlying contour tones. *Working Papers in Phonetics 27*, (Los Angeles: University of California), 74-83.

Erickson, D. M. (1974) Fundamental frequency contours of the tones of standard Thai. *Pasaa 4*, 1-25.

Erickson, D. M. (1976) *A Physiological Analysis of the Tones of Thai.* (Ph.D. Dissertation, University of Connecticut).

Gandour, J. (1974) On the representation of tone in Siamese. In *Studies in Thai Linguistics in Honor of William J. Gedney*, ed. by J. J. Harris and J. R. Chamberlain. (Bangkok: Central Institute of English Language), 170-195.

Leben, W. R. (1973) The role of tone in segmental phonology. In *Consonant Types and Tone*, ed. by L. M. Hyman. (*Southern California Occasional Papers in Linguistics 1*), 115-149.

Sarawit, M. (1976) Assimilation in child language acquisition. In *Tai Linguistics in Honor of Fang-Kuei Li*, ed. by Thomas W. Gething, Jimmy G. Harris, and Pranee Kullavanijaya. (Bangkok: Chulalongkorn University Press), 205-213.

Trager, G. L. (1957) Siamese phonemes: A restatement. In *Studies Presented to Yuen Ren Chao on his Sixty-Fifth Birthday*. (Taipai: Academia Sinica).

Closure Hiatus: Cue to Voicing, Manner and Place of Consonant Occlusion[*]

Leigh Lisker[†]

ABSTRACT

Delayed onset of laryngeal vibration following release of an initial stop by about 35 ± 15 msec generates acoustic features eliciting p,t,k responses from speakers of English. These features, by a common misnomer, are referred to as cues to stop voicelessness; in fact, they are cues to voiceless stop aspiration. Medially, before unstressed vowels, English has voiceless stops that are not aspirated, and these lack some of the features of initial /p,t,k/. An important cue to medial /p,k/ before unstressed vowels is an interruption of glottal pulsing during closure, provided this interruption exceeds a certain duration. In experiments replicating and extending earlier studies, a number of naturally produced and synthesized polysyllables were varied in respect to their closure intervals. In part, results replicated earlier findings, but not unambiguously. It appeared that 1) there were significant individual differences in response to stimuli with edited closure intervals; 2) stimuli derived from different tokens of the same phonetic types elicited different responses; 3) the apical flap ([ɾ]) response to very short closure intervals could not be entirely explained by a simple motor theory interpretation.

The recent literature dealing with acoustic cues that separate homorganic stops in English is mostly concerned with stops initially before stressed vowels. With respect to the most important of these--the time of onset of laryngeal pulsing--we are told that /ptk/ is distinguished from /bdg/ along this so-called VOT continuum, in that for /ptk/, the onset of pulsing must be deferred either until a certain time, about 35 msec, after the stop release, or until the articulatory shift from closure to succeeding vowel has been largely completed. The fact that this requirement for initial /ptk/ cannot hold true for phonetic events--linguistically identified with /ptk/--that occur in other contexts, has been relatively unemphasized. If, for example, we say that the word paper includes two instances of /p/, the VOT requirement

---

[*] This paper was presented orally at the 93rd meeting of the Acoustical Society of America, Pennsylvania State University, 5-8 June, 1977.

[†] Also University of Pennsylvania.

Acknowledgment: Acknowledgment is made of the support of the National Institute of Child Health and Human Development.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 1 (1978)]

just mentioned must be satisfied only for the initial one, since the medial /p/ does not usually involve much of an interval between release and resumption of pulsing. The acoustic properties of the medial /p/ following its release are in fact usually of a kind that will elicit *b* judgments, if the signal in which the /p/ is embedded is edited so that the release burst and transition come to be in initial position. Given this fact, how can it be said that any acoustic feature depending on a delay in pulsing onset is *the* cue to stop voicelessness in English? All that can be claimed is that the presence of such a feature is sufficient to trigger *ptk* responses; in its absence *bdg* responses are not necessarily reported.

The attention lavished on the VOT dimension, and the importance attached to particular durations by which pulsing onset lags behind release, reflect the fact that so much of the search for the segmental cues has been focused on the analysis and synthesis of nursery utterances such as *ba da ga*. This is despite the fact that in a piece of speech all but one phonetic event is noninitial, and it is not generally believed that speech is made up entirely of simple concatenations of CV sequences like those that can occur as complete utterances. For the stops even more than for other classes of phonetic events, at least in English, it is a mistake to gloss over the context-dependent nature of the cues by which /ptk/ and /bdg/ are distinguished. If primary attention had been directed to medial or final position, we should have a somewhat different idea of the acoustic basis for the distinction. Because both /ptk/ and /bdg/ occur initially and both can occur finally without any acoustic signal of release, it would seem impossible to claim that any feature found either before or after closure is a necessary property for the perception of either class of phonemes.

Leaving aside the case of the final stops, let us consider some evidence that an intervocalic occlusion with interruption of pulsing may be interpreted as either /bdg/ or /ptk/ when pulsing resumes immediately upon the release. This evidence comes from an old experiment, (Lisker 1957), since replicated in some recent work by Robert Port (1976), that involved the editing of natural speech recordings so as to vary the duration of a silent interval corresponding to an intervocalic closure. Manipulation of tokens of the words *ruby* and *rupee* yielded stimuli that a group of seven phonetically naive subjects labeled as shown in Figure 1. It appears that the duration of silent closure may figure as a significant cue for word identification, specifically for the /p/-/b/ contrast. The *rupee*-derived stimuli were heard mostly as *ruby* for closure durations less than 70 msec; *ruby*, when its buzzed closure was replaced by silence longer than 100 msec, was more often reported as *rupee*. (The two intermediate curves of the display, giving responses to stimuli composed of cross-combinations of first and second syllables of the source words, will not be commented on now.) The difference in cross-over values for the *ruby* and *rupee* curves is a little more than 30 msec, and we may like to think of this difference as the perceptual-phonetic equivalent of whatever other features that precede and follow the silent interval and also operate as cues.

The range of durations tested in this experiment was chosen with a lower limit of 40 msec so as to exclude the possibility that listeners would report hearing a *t* or *d* (that is, the alveolar flap) rather than *b*, while the upper bound of 140 msec was intended to avoid the effect of an abnormally long or

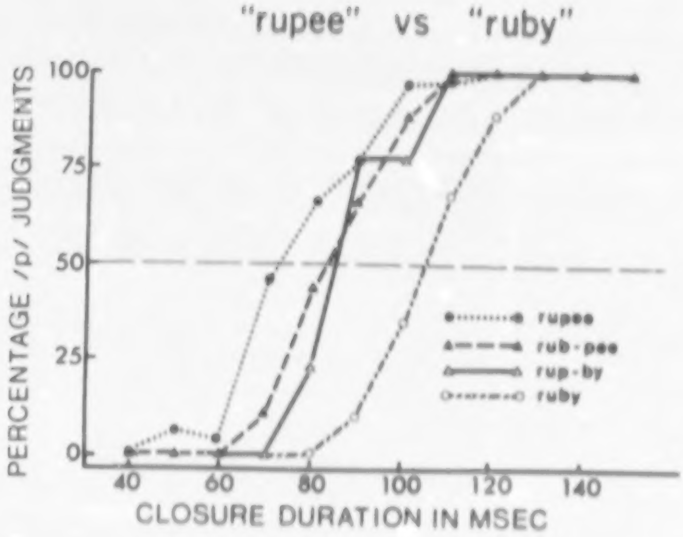


Figure 1: Labeling responses of seven naive listeners to 44 stimuli derived by tapecutting recordings of naturally produced tokens of the target words, and recombining pre- and post-closure intervals in four ways, with silent intervals varying in 10 msec steps from 40 to 140 msec.

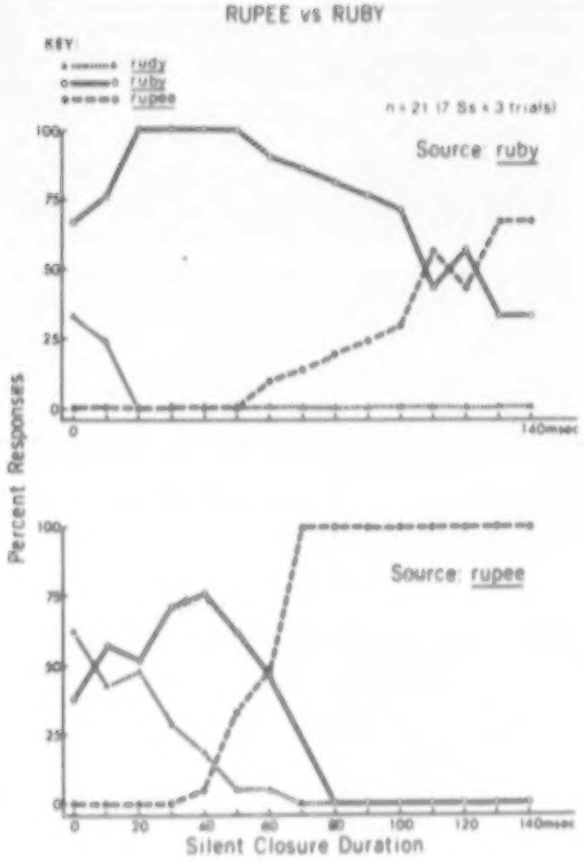


Figure 2: Labeling responses of seven listeners to spoken ruby and rupee edited so stop closures of each word were silent intervals varying in duration from 0 to 140 msec, in 10 msec steps. Subjects were instructed to listen for rupee ruby Rudy.

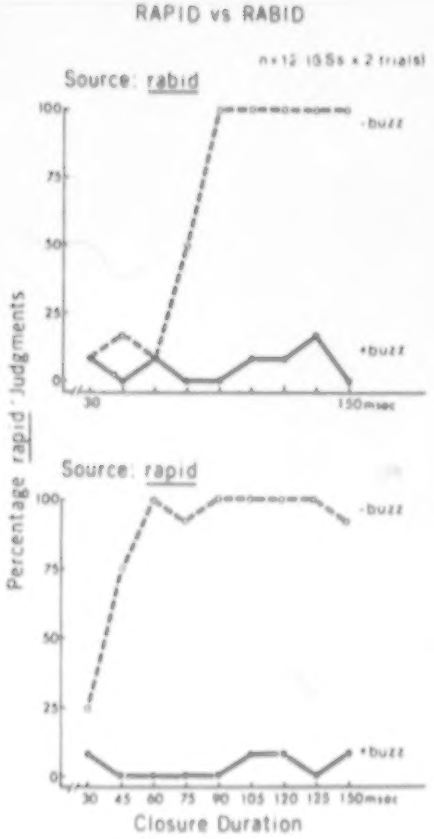


Figure 3: Labelings of edited natural productions of rabid and rapid. Closure intervals, varying in 15 msec steps from 30 to 150 msec, were either silent or filled with naturally produced glottal buzz. The twelve listeners were phonetically naive.

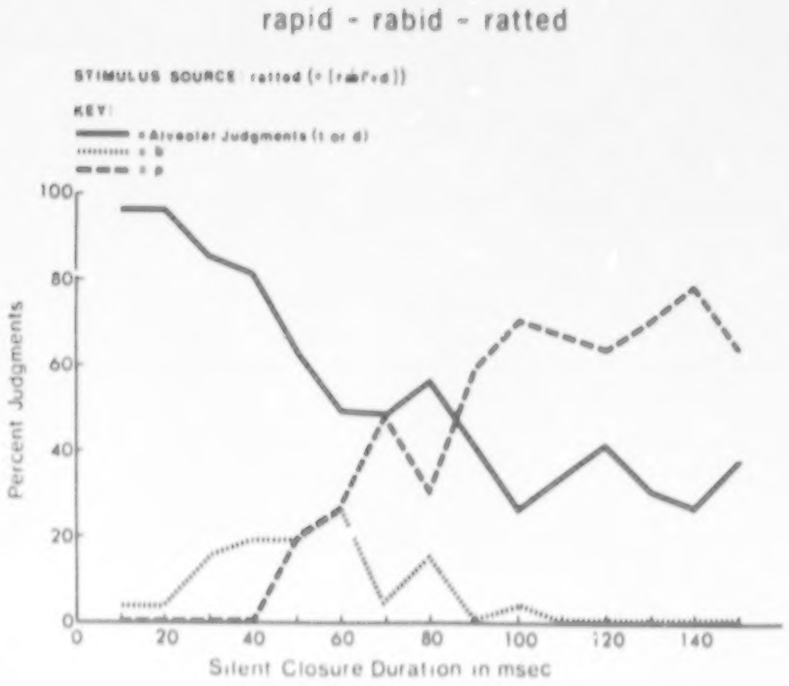


Figure 6: Responses of 15 naive listeners to stimuli derived from a naturally produced token of ratted ([rǽɾɨd]). A silent "closure" was introduced at the point where waveform and spectrogram indicated maximum constriction. Subjects were allowed free choice of responses.

geminate *p*. The shift in place and manner judgments at closure of less than 30 msec has been studied in detail by Port, (1976). The conclusion is reasonable that closure duration not only may serve as a cue to stop voicing, but that it must have some minimum duration if it is to signal a stop consonant. At durations too small to be appropriate for the perception of stop manner (and probably for stop production as well), the consonant perceived was a flap. Whether it is because the only flap in English is apico-alveolar, or because there is some purely auditory basis for the perceptual zeroing of the labial place cues, listeners often reported hearing *rudy* instead of *ruby*. While it might be more interesting, especially to the linguist, if the first account were true, I think the second is closer to the mark. The basis for this belief will be made clear later on.

Now I want to turn to some recent experiments, first of all to one performed to see whether the old results would be replicated. Figure 2 represents labeling responses for silent durations ranging from 0 to 140 msec, the upper panel for stimuli derived from a token of *ruby*, and the lower for those from *rupee*. The *ruby*-derivatives were heard mostly as *ruby* for closures of from 0 to 100 msec. Stimuli derived from *rupee* were heard as *rupee* for durations of 70 msec and greater. The /b/-/p/ crossover values are not very different from those measured in the first experiment. Unlike the older results, here no *ruby*-derivatives achieved better than 70 percent *p* responses, while no stimuli from *rupee* were reported as *ruby* more than 75 percent of the trials. In the case of the flap, there is no duration for which this category included more than 60 percent of the responses, a score achieved only for a *rupee* with closure interval reduced to 0 msec duration. I think this result does not mean that Port's somewhat different findings cannot be replicated--only that it cannot be guaranteed that every token of an intervocalic /b/ or /p/ can be heard as the flap for very brief closures, even those too short for human vocal tracts to execute.

Figure 3 shows responses to a set of stimuli derived from another favorite pair of words: *rabid* and *rapid*. In this set, the closure duration was varied in 15 msec steps, from 30 to 150 msec. For each value of closure duration, two stimuli were prepared, one silent and the other filled with buzz of laryngeal origin. The dashed lines are for the buzz-filled. At least for the particular token of *rabid* used, *rabid* → *rapid* for silent closures of 90 msec and more, a shift occurs that is much more decisive and at a smaller crossover interval than in the case of the second *ruby*-*rupee* test. For the stimuli derived from *rapid*, the shortest silent closure rated no better than 75 percent of the *b* responses. For both *rabid*- and *rapid*-derived stimuli, the introduction of buzz into the closure interval shifted judgments decisively--80 percent or more--to *b*. None of the six phonetically untrained listeners reported anything other than *rabid* or *rapid*. By contrast, Figure 4 shows labeling responses of three trained listeners. Responses to *rabid*-derivatives with silent closures, shown in the lower left, fall into five categories: flap, *b*, geminate *b* (*bb*), *p* and geminate *p* (*pp*). With 30 msec of closure, all responses report the flap category, *b* responses are at a maximum (but no more than 50 percent) for a silent interval of 60 msec, while for 75 msec and longer most responses are *p* or geminate *p* (*pp*). Responses to the stimuli derived from *rapid* and having silent closures are shown in the lower right. The *b* responses are even fewer than in the case of the *rabid*-derived items, and *p* responses preponderate, starting with a closure of 60 msec. For the shortest closures the flap responses are also fewer than those elicited by the stimuli having *rabid* as their source. The upper panel, giving responses to the stimuli with buzz-filled closures, shows only three categories: flap, *b* and geminate *b* (*bb*). Here, too, the stimuli from *rabid* were most often reported as *ratted* for the shortest closures. It should be remembered, however, that this is not in agreement with the finding for *ruby-rupee*, where it was *rupee* whose derivatives were more often heard as the form with a medial apico-alveolar flap.

The next experiment was undertaken to discover whether the finding for *ruby* vs. *rupee* and *rabid* vs. *rapid* meant that silent closure could be said to operate as a cue to stop voicing independent of place of articulation. The word-pair used was *locker-lager*, which in my variety of American English differ only with respect to their stop consonants. Figure 5, in the top panel, suggests a strong place effect: the 0 interval between end of implosive transition and release burst reduces the *locker* token to something ambiguous between *locker* and *lager*, while increasing the gap to the longest one tested is no more effective in shifting *lager* to *locker*. The subjects were instructed to listen for a nonsense form *latter* ([láɾɚ]), but reported only *locker* and *lager*. The pooled data of the upper panel, when examined for individual patterns of labeling behavior, revealed that the subjects could be divided, nonarbitrarily, into two groups of three each. Group 1 reported all *locker*-derivatives as *locker* in more than 75 percent of their responses, even for closure intervals of 0 and 10 msec; moreover *locker* was reported for the *lager*-derived stimulus with the longest silent gap. Group 2 showed a bias the other way: *locker* went to *lager* for the shortest closures, but *lager* remained *lager* 100 percent independent of the closure duration. None of the six subjects was prepared to accept both a shortened /k/ closure as /g/ and an augmented and silenced /g/ closure as /k/. No doubt we can, by pure synthesis, tailor stimuli of a kind to enhance the effectiveness of silent interval as the feature controlling a shift between medial *g* and *k*, but the present data cast some doubt on the view that closure duration *per se* functions crucially in natural speech. If, in our figure, we restrict our attention to the range of closure durations recorded in natural speech, say from 30 to 100 msec (see, for example Sharf, 1964), then none of the curves in the display crosses the 50 percent level.

The last experiment is one in which a dissyllable *ratted* (past tense of the verb "to rat") was edited in the usual way and submitted to listeners for labeling. Their responses, shown in Figure 6, are rather surprising; as the silent interval increases, there is a shift from mainly *t* or *d* to mainly *p* judgments, with *b* not exceeding 27 percent. These data do not tell us where *t* + *d* judgments represent flap percepts and where they represent stops (since the subjects had not had enough phonetic experience to make such a discrimination). However, the fact that labials were reported, together with the fact that labials, but not velars, could be converted to flaps, suggests that the place information generated by the apico-alveolar flap articulation is ambiguous, and that this ambiguity has some acoustic basis other than a simple temporal one. Casual inspection of some spectrograms does not make this seem unreasonable. An explanation for the shift from labial to apico-alveolar flap (or simply *t* to *d*) judgments that appeals to the fact that only at the latter place can we produce closures of 30 msec and less cannot be turned around, for we cannot claim that closures of 90 msec and more can be produced only at the bilabial place of articulation.

To summarize: 1) there are significant differences among subjects to the extent in which their labelings of silent closure intervals as /ptk/ or /bdg/ are duration-controlled; 2) response patterns differ, in crossover values and cleanness of category separation, when different tokens of the same words serve as stimulus sources; 3) if we consider the two places of articulation where stops are produced in trochaic words in American English, labial and velar, and particularly if we limit attention to closure durations commonly found in speech, then the nature of the closure interval, silent vs. buzz-filled, seems a more reliable predictor of labeling behavior than does the duration of that interval; 4) the perception of labially produced closures as alveolar flaps when the durations are very short depends, at least partially, on the failure of alveolar flap articulations to produce place cues clearly distinguishable from those produced by bilabial stop articulation.

REFERENCES

Lisker, L. (1957) Closure duration and the intervocalic voiced-voicelessness distinction in English. Language 33, 42-49.

Port, R. (1976) The influence of tempo on the closure interval cue to the voicing and place of intervocalic stops. Journal of the Acoustical Society of America 59, Suppl. 1, S41-42.

Sharf, D. J. (1964) Duration of post-stress intervocalic stops and preceding vowels. Language & Speech 5, 26-30.

Metaphoric Comprehension: Studies in Reminding and Resembling

Robert R. Verbrugge[†] and Nancy S. McCarrell[††]

ABSTRACT

The theoretical problems posed by metaphoric comprehension are discussed in the context of experiments on prompted recall. Listeners heard sentences of the form "Topic is (like) Vehicle." In most cases, a statement of the implicit resemblance (the "ground") was very effective in prompting recall of its related metaphor. This result could not be attributed to the activation, transfer, or additive combination of pre-existing properties of the topic and vehicle terms or to pre-existing associations between grounds and sentence terms. It is argued that the vehicle domain guides a novel schematization of the topic domain, that the perceived resemblance is a higher-order relation among entities (both explicit and implicit) in each domain, and that this abstract relation constitutes the "functional memory unit." Prompted recall may begin with recognition of this previously experienced relation.

[†] Also University of Connecticut, Storrs.

[††] George Peabody College for Teachers, Nashville, Tennessee.

Acknowledgment: This research was supported in part by a fellowship to Robert Verbrugge from the University of Michigan Society of Fellows; by grants to the University of Minnesota, Center for Research in Human Learning, from the National Science Foundation (GB-35703X), the National Institute of Child Health and Human Development (HD-01136), and the Graduate School of the University of Minnesota; and by a grant (HD-00973) from the National Institute of Child Health and Human Development to George Peabody College for Teachers. We wish to thank John Bransford, James Jenkins, Dane Harwood, James Reaves, Robert Shaw, and Buford Wilson for their advice and encouragement throughout this research and Mark Jaffee for his assistance in data analysis. A preliminary report on Experiments I and II was made to the Midwestern Psychological Association, Chicago, May 1973. The research was undertaken while both authors were at the Center for Research in Human Learning (University of Minnesota). The Human Performance Center (University of Michigan) and Haskins Laboratories provided assistance during preparation of the manuscript and this support is gratefully acknowledged.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 1 (1978)]

Metaphoric language invites a "perception of resemblances,"[1] and the invitations come in many forms. Examples of metaphors (strictly defined) are these: My lawnmower is a wild animal; The children galloped to the cafeteria; Billboards are warts on the landscape. In these cases, a resemblance is communicated by forms that assert or presuppose an identity. Similes and analogies are less bold since they directly assert a relation of similarity: The freeway is like a snake; He runs as fast as a cheetah. Beyond these, there are dozens more "hedge" forms by which similarities are expressed (Lakoff, 1972); for example, George resembles a truck driver; Judy is kind of a prima donna. In each of these metaphoric forms, two domains are being compared: a topic (traditionally called the tenor; Richards, 1936) and a vehicle (that to which the topic is being compared). The topic is usually mentioned explicitly, but in such forms as proverbs, parables, and allegories it must be supplied by the comprehender. Similarly, the vehicle may be mentioned explicitly or it may simply be alluded to, as in the galloping sentence above. The resemblance between the topic and vehicle domains is traditionally called the ground (or tertium comparationis). The ground is occasionally made explicit (as in the cheetah sentence above), but usually it is the reader's or listener's task to discern the resemblance. The task for psychologists, in turn, is to characterize the structure of the apprehended resemblance, its relationship to the terms that appear in a sentence, and the process by which the resemblance is discerned.

Psychologists and linguists have devoted comparatively little attention to the meaning and comprehension of metaphoric language. Part of the explanation may lie in the long tradition in epistemology and rhetoric that stresses the categorization of reality in terms of elementary sensory or semantic features, the sharply defined and enduring character of these features, and the relative stability of their interrelations. If such a semantics is presupposed, metaphor can pose a special problem for explanation, since it often demands that we accept a categorization radically different from what is familiar or conventional. It is a short step to viewing metaphor as an illogical and even freakish language form--an object of universal fascination, perhaps, but one that resides at the periphery of ordinary language use. This academic attitude ignores what seems obvious to casual observation: metaphoric language is endemic to ordinary communication. It is common in day-to-day conversation, narrative, popular songs, newspaper articles, effective teaching, and problem solving. In fact, metaphor may be basic to all growth in understanding, whether in the playroom, the classroom, the psychotherapeutic setting, the scientific laboratory, or the theater (see Hesse, 1966; Langer, 1967; Verbrugge, 1977; Pollio, Barlow, Fine and Pollio, 1977).

Since appreciation of the importance of metaphor has developed only recently in psychology, research on metaphoric comprehension (particularly in adults) has been sparse. Though the research is difficult to classify systematically, it is convenient to identify two traditions: associationism and transformational linguistics.

---

[1] We have borrowed this phrase from Aristotle, whose views on poetic language are expressed in his Poetics and Rhetoric. A summary may be found in Hawkes (1972).

Associationism proposes that words are associated with an array of elemental ideas, concepts, images, and combinations thereof, and that a probability or strength can be assigned to each of these links. Sentence meaning is some kind of composite of the associations to constituent elements. Metaphors are viewed as fortuitous, low-probability associations, governed by the usual laws of conditioning and transfer. One option is to view the topic and vehicle as having common associates: words with "stimulus equivalence" are linked when producing the metaphor, and comprehenion involves activating the common associate (see Asch, 1955; Skinner, 1957). A related view is that metaphor involves the substitution of a response for one that is more typical and appropriate (Osgood, Suci and Tannenbaum, 1957; Brown, Leiter and Hildum, 1957; Koen, 1965). For example, *The baritone's voice was heavy* might be spoken in response to hearing a singer's voice, due to the strong associations between *low-pitched voice*, *large body*, *heavy*, *loud*, etc., in prior experience. Comprehension involves activating these high-frequency ("literal") associates and linking them to the topic. While the more sophisticated theories of associative networks (for example, Anderson and Bower, 1973; Norman and Rumelhart, 1975) have seldom been applied to metaphoric sentences, they propose representational structures and procedures that are similar to those just described. Comprehension of a metaphoric sentence would presumably involve detecting common associated predicates in the network or transferring predicates from one node to another.

A second influential approach to the psychology of metaphor is an outgrowth of transformational linguistics. In the semantic systems proposed by Katz and Fodor (1963) and Chomsky (1965), sentence constituents were indexed in a lexicon by grammatical category, a set of distinctive semantic features, and selection restrictions that defined the contexts in which a term could appear. Expressions that failed to honor these restrictions were labeled semantically unacceptable, anomalous, and deviant. Among this riff-raff of rejected word strings were many varieties of figurative language, including metaphor. Other linguists, not wishing to lose metaphor as an object of linguistic description, have suggested that special rules be added to a grammar to permit interpretation of these "deviant" sentence forms (for example, Weinreich, 1966; Bickerton, 1969; Leech, 1969; Matthews, 1971). Common strategies have been to suspend selection restrictions temporarily, to ignore incompatible feature values, or to alter the standard feature descriptions for terms (for example, by reassigning values to some of their features). These are temporary alterations to the language device, allowing it to process abnormal inputs that would otherwise bring it to a grinding halt. To a large extent, these efforts have shared the basic assumptions of the Katz and Fodor (1963) model: metaphor is a semantic violation; its identity and interpretation are to be characterized without reference to the intentionality, nonlinguistic knowledge, or processing strategies of language users; and the special rules operate on stable semantic feature descriptions associated with terms. On the latter point, this approach is similar to traditional associationism, except that a highly constrained structural organization of features is proposed.

The linguistic approach to metaphor sharply distinguishes between sentences that are well formed and anomalous, normal and deviant, acceptable and unacceptable. Many psychologists of language have accepted this dichotomy, focusing their research on "rule-governed" language and contrasting its processing with that of "anomalous" language (for example, Marks and Miller, 1964; Steinberg, 1970; Epstein, 1972; Collins and Quillian, 1972; Smith, Shoben and Rips, 1974). In the few cases in which metaphoric "anomaly" has been the focus of psychological research, the characterization of meaning is similar to that found in associative accounts. For example, Johnson, Malgady and Anderson[2] and Malgady and Johnson (1976) have attempted to define the operations on two partially incompatible feature sets that could yield the appropriate ground (that is, common associated features) as a product. Kintsch (1972, 1974) has argued that metaphors are anomalous surface forms produced by condensation of deep-structure assertions of similarity. In this model, certain "lexical implications" and properties are already associated with both the topic and the vehicle, and comprehension includes a search for associations shared by the two terms.

It is important to determine why associative and linguistic models have shown only localized and transitory success as theories of metaphoric language. We believe that two important hindrances to success have been the following.

(1) Metaphor has been treated as uniquely ambiguous, imprecise, and illogical. In most logics of this century (including that underlying semantic feature theory), meaning is assumed to be sharply bounded, that is, the criteria for ostensive application of a term to a referent are (in principle) precisely and unambiguously defined. Imprecision in language use is attributed to the difficulty encountered by a speaker-hearer in relating the criteria to a specific situation, that is, it is a "performance" phenomenon. Verbal ambiguity, therefore, could result from poor viewing conditions, inattention, carelessness, immaturity, or psychopathology. If metaphor is viewed as an imprecise application of terms to referents, it is a short step to interpreting the metaphoric productions of adults, children, schizophrenics, and poets as "deviant." But more important than this invidious labeling is the conclusion drawn about comprehension: to understand the anomaly one must rationalize it according to the sharply defined constraints that apply to ordinary language. Accordingly, most of the recent accounts of comprehension assume that the listener must "normalize" a metaphor, that is, intuit the literal (precise) meaning that must have been intended.

Dissatisfaction with this view of meaning criteria has grown in recent years. An increasing number of linguists and psychologists have come to believe that semantic feature classification is inadequate for explaining the flexibility and precision of ordinary language (for example, Bolinger, 1965; Cohen and Margalit, 1972; Rosch, 1973; Anderson and Ortony, 1975; Bransford,

---

[2] Johnson, M. G., Malgady, R. G. and Anderson, S. J. Some cognitive aspects of metaphor interpretation. Paper presented at the meeting of the Psychonomic Society, Boston, November 1974.

McCarrell and Nitsch, 1976). One leitmotiv in this dissent is the belief that the underlying criteria for word use are not sharply defined; they are "fuzzy" and abstract constraints. One goal of current theoretical efforts is to understand how *precision* may be achieved by the application of the constraints in particular contexts (see Lakoff, 1972; Bransford and McCarrell, 1974). If the standard uses of terms are only fuzzily bounded, the distinction between metaphoric and literal language itself becomes fuzzy, and the goal of rationalizing one in terms of the other becomes suspect. In a fuzzy logic, the use of a term is always metaphorical in the following sense: a new context of use has only a *sufficient resemblance* to prior contexts of use. If we say *This penguin is a bird* or *This creature is a penguin*, we are making an assertion about a sufficient resemblance to prototypical constraints on birdiness or penguinicity. The process is very similar when we say *My daughter is a bird* or *That cloud is a penguin*; again, these are motivated by the applicability of a set of abstract constraints to a novel instance. Thus, the apparent precision and primacy of literal language dissolves when we realize that all language use occurs in novel contexts, and that these contexts are related by a *sufficient resemblance*, not an *identity* defined by invariant criterial features. Metaphoric and literal assertions seem to part company over how exhaustively the conventional constraints apply, not in precision. (Compare a penguin-shaped cloud, a portly gentleman in a tuxedo, and a real penguin.)

(2) A second major hindrance to success in developing a theory of metaphor has been the characterization of grounds in terms of common features and common associations. Metaphoric comprehension has been treated as a kind of concept formation task in which the concepts are "attributive," that is, word meaning is defined by a set of associated attributes. The process is one of "subtractive" concept formation, since shared attributes become part of the ground, while conflicting attributes are ignored. The attributes (features, properties) are treated as substantive building blocks of identity, both in the narrow sense of linguistic meaning (they are elements that concatenate to form word meaning) and the broader sense of knowledge about the referent (they are elements that concatenate to form factual knowledge). The underlying theoretical metaphor has changed little through the long history of associationism: attributes are substantive atoms.

We need to consider carefully whether attributive concepts are sufficient to characterize the grounds of metaphors. Many metaphors draw attention to common systems of relationships or common transformations, in which the identity of the participants is secondary. For example, consider the sentences: *A car is like an animal*, *Tree trunks are straws for thirsty leaves and branches*. The first sentence directs attention to systems of relationships among energy consumption, respiration, self-induced motion, sensory systems, and, possibly, a homunculus. In the second sentence, the resemblance is a more constrained type of transformation: suction of fluid through a vertically oriented cylindrical space from a source of fluid to a destination. In each case, the substantive components of the two domains show little or no resemblance. Translating the relationships into attribute lists is an awkward and unbounded process and may be impossible in principle.

There have been many efforts to characterize such systems of relationships or "schemata," to distinguish them from attributive concepts, and to argue against the adequacy of attributive concepts as the primary basis for conceptual knowledge (for example, Cassirer, 1923; Piaget, 1950; Jenkins, 1966; Bransford and Franks, 1973; Weimer, 1973). For present purposes, we will speak of these relational systems as abstract relations, to emphasize that the structure of resemblance is primarily abstract.

A particularly useful characterization of such relations is found in the discussion of event perception by Shaw, McIntyre and Mace (1974). These authors characterize an event in terms of a *transformational invariant* (a kind of transformation exerted over a structure, for example, rotation) and a *structural invariant* (what the transformation leaves invariant, for example, spherical shape). Either type of invariant or both can serve as the basis for a resemblance. For example, in the *tree trunk* sentence, the flow of fluid is a transformational resemblance: the transformation leaves the tubular structure and the volume of fluid invariant in each domain. Since both the tree trunk and the straw have a tubular structure, this constitutes a structural resemblance that enhances the strength of the metaphor. It is tempting to view the structural resemblances as attributes of the traditional kind. It is important to keep in mind, however, that such invariants always presuppose some transformation or system of relationships, and that these are contextually variant. Thus, in *Tree trunks are pillars for a roof of leaves and branches*, the structural invariant is a solid column rather than a hollow tube. The tree trunk is not the same "structure" in each case; for this reason, a fixed set of properties could not characterize its role in the two different metaphors. In general, attributive concepts fail by overlooking transformational resemblances, by assuming that the resemblances draw on a fixed, contextually invariant set of structural primitives, and by assuming that structural primitives are substantive in kind (rather than abstract or mathematical).

The research reported here focuses on the structure of metaphoric resemblances. Identifying the structure of grounds is a crucial prerequisite to studying how they are discerned. Their structure places important constraints (and demands) on the class of process models one might consider. Traditional definitions of the ground in terms of shared attributes led naturally to models involving feature search, comparison, weighting, and transfer. It is important to determine whether features associated with the nominal terms (objects) in a metaphor are an adequate basis for defining the resemblance discovered by the ordinary listener. The event or relationship in which the objects participate may be more critical in defining the resemblance. If so, a different class of comprehension models is necessitated, in which, for example, salient transformations over the vehicle domain are applied over the topic domain.

The accessibility of acquisition material to recall can provide a sensitive symptom of how the material was interpreted. It is becoming increasingly clear that a person's "orienting task" (whether adopted autonomously or at the experimenter's request) has as distinctive an effect on recall as the properties of the materials themselves (see Jenkins, 1974; Craik and Tulving, 1975). Prompted recall is especially useful as a measure of comprehension, since it is differentially sensitive to components that are *central* to sentence meaning (Blumenthal, 1967; Blumenthal and Boakes, 1967; Perfetti and Goldman, 1974), and it is sensitive to information supplied implicitly by the comprehender (Tulving and Thomson, 1973; Barclay, Bransford, Franks, McCarrell and Nitsch, 1974; Anderson and Ortony, 1975).

In the case of metaphoric sentences, prompted recall may provide a sensitive measure of the presence of inferential activity during comprehension, the kind of resemblances inferred, and the context specificity of a topic's interpretation in different metaphors. Specifically: (a) If an abstract relation is central to what is comprehended from a metaphor, a verbal precis of the relation should be an effective prompt for the sentence's recall (even if no terms in the precis match terms in the original sentence). Abstract resemblances of this sort have proven to be effective prompts for recall of proverbs (Bühler, 1908; Honeck, Reichmann and Hoffman, 1975). (b) If the topic is interpreted uniquely in different metaphors (for example, as a participant in different types of events), then a possible "ground" should only prompt recall of the topic when it specifies the relevant type of event or relationship. For example, the ground for the *tree trunks-straws* metaphor might be summarized verbally as follows: *are tubes which conduct water to where it's needed*. This phrase might effectively prompt recall when tree trunks have been compared to straws, but it may not be effective when tree trunks have been compared to pillars, even though it expresses a perfectly valid property of tree trunks. A more effective prompt for the *tree trunks-pillars* metaphor might be *provide support for something above them*, since it expresses the resemblance which is specific to the *pillars* context of interpretation. By using pairs of acquisition lists with common topics and prompting recall with possible grounds, one can test whether such specific interpretations are made. Previous studies on prompted recall have demonstrated this kind of "encoding specificity' for terms in literal sentences and word lists (for example, Thomson and Tulving, 1970; Anderson and Ortony, 1975).

In the experiments reported here, we have studied metaphors that are expressed linguistically, explicitly, and in sentence form, that is, cases where a perceived resemblance is communicated through words, where both the topic and the vehicle are explicitly mentioned, and where the comparison is made within a single sentence rather than in a text or a discourse of greater length. We have used two sentence forms, metaphor ("A is/are B") and simile ("A is/are like B"), and the grounds are combinations of both transformational and structural resemblances. Hypotheses based on abstract relations will be tested in parallel with a series of recall models framed in the language of features. The effort throughout this study is to identify the structure of the comprehended resemblance and its relationship to the terms in a metaphoric sentence.

EXPERIMENT I

This experiment tested whether the ground of a metaphor can be an effective prompt for its recall. The design of the study crossed two acquisition lists (with matched sets of topics) with two sets of recall prompts. Subjects received ground prompts that were all relevant or all irrelevant to the original list of metaphoric sentences.

The rationale for this design was as follows: if a verbal statement of the ground successfully prompts subjects' recall, one might challenge the conclusion that the ground had been inferred during an acquisition process guided by the vehicle. Since the ground states a property that is true of the topic, it might serve as an effective prompt irrelevant of any special interpretation guided by the vehicle. Semantic network and semantic feature theories both suggest that major constituents of a sentence independently activate an array of associated predicates or attributes. Thus, the ground in question may be activated whenever the topic appears. For example, both *are tubes which conduct water to where it's needed* and *provide support for something above them* may be activated in response to *either* acquisition sentence about tree trunks and, therefore, might appear in the record of either event. Alternatively, acquisition sentences could be stored more-or-less verbatim, and the subject's strategy at recall could be to scan this record for a topic that contains the ground in its feature list or for which a path to the ground can be found in the network.

To control for these possibilities, two kinds of prompts may be used: (a) a set of "relevant grounds" in which each prompt is relevant to the sense of an acquisition metaphor, or (b) a set of "irrelevant grounds" in which each prompt is irrelevant to the sense of a particular metaphor, but is nonetheless true of its topic. If the vehicle does affect interpretation of the topic, relevant grounds should be more effective as prompts than irrelevant grounds. To insure that this difference between the two sets of grounds is not artifactual, one can present a second group of subjects with a list of acquisition metaphors (using the same topics) for which the formerly "relevant" grounds are now irrelevant and the formerly "irrelevant" grounds are now relevant. The ordering of prompt effectiveness should reverse, even though the topics involved are the same in both cases.

Method

Materials. Two lists of 14 metaphoric sentences were prepared (lists A and B). The topics in each list were the same, while the vehicles were different. For example, tree trunks were compared to pillars in List A and to straws in List B. The various topics and vehicles were kept as distinct as possible; with the exception of the paired topics, no nouns or close synonyms were repeated elsewhere in either list. This was intended to minimize systematic errors in recall. The lists were recorded on audio tape by an adult male speaker using a natural speaking pace, amplitude, and intonation contour. Each sentence was spoken twice. There was a 5-sec pause between the repetition and the next sentence in the list. Topics appeared in the same order in each list.

A "ground" was prepared for each of the 28 metaphoric sentences for use as a prompt. The ground took the form of a predicate expression. It was intended to summarize the major resemblance underlying the metaphor, but was not assumed to be an exhaustive interpretation. The following are further examples of acquisition sentences (and grounds) used in the study.

Skyscrapers are honeycombs of glass. (are partitioned into hundreds of small units)
Skyscrapers are the giraffes of a city. (are very tall compared to surrounding things)
Billboards are warts on the landscape. (are ugly protrusions on a surface)
Billboards are the yellow pages of a highway. (tell you where to find businesses in the area).

The acquisition sentences were written to keep the 28 grounds as dissimilar as possible, again to avoid systematic intrusions in recall. In particular, the pairs of grounds for each topic were chosen to be as unrelated to each other as possible. Each ground avoided content words appearing in the related sentences and terms that are typically constrained to either the topic or vehicle context.

The grounds were assembled into two sets of prompts, grounds A and B, for use in recall. Grounds A were the 14 grounds relevant to the sentences in List A; Grounds B were relevant to List B. The grounds were typed on individual slips of paper, with ample space for subjects to write out a sentence during recall. Each set of prompts was presented in booklet form; a blank slip of paper on top of the booklet obscured the first prompt from view.

In addition, prompt booklets containing the topics and vehicles from the original sentences were prepared. Topics A and B were identical and contained the full-subject noun phrases from the 14 sentence pairs. Vehicles A and Vehicles [illegible] contained the vehicles from the related acquisition lists. In some cases the full predicate noun phrase was not included. If a word or phrase in the predicate (for example, *leaves and branches*) was related to the topic domain (tree trunk), it was excluded from the vehicle prompt.

The order of prompts in all booklets was randomized with respect to the acquisition order, and the *same* order was used in all cases (that is, the order of correct recall would be the same).

Subjects. Subjects were 96 undergraduates enrolled in an introductory psychology course at the University of Minnesota. They received extra credit for their participation. Subjects were randomly assigned to one of two list conditions, List A or List B. In each condition, 8 subjects received the related topic prompts (Topics A or B), 8 received the related vehicles (Vehicles A or B), 16 received Grounds A, and 16 received Grounds B.

Procedure. In each session a group of subjects sat in a small experimental room facing a tape recorder placed on a desk at the front. The experimenter informed them that they would hear a series of metaphoric sentences describing various types of people, emotions, objects, and so on. They were asked to listen to the sentences and think about what each one was trying to express. No mention was made of a subsequent recall task. After playing List A or B, the experimenter informed subjects that they would receive a booklet containing phrases related to the sentences they had just heard. They were asked to write out the full sentence that each phrase reminded them of most. The experimenter then distributed the prompt booklets and paced the subjects at 40 sec per prompt.

Results

Sentences were scored correct if a subject recalled both the topic and the vehicle. A topic or vehicle was considered correct if it included the central noun from the original topic or vehicle noun phrase. Paraphrases were accepted if close synonyms were substituted for topic or vehicle terms and if the order of topic and vehicle was reversed.

The mean proportion of sentences correctly recalled by subjects is recorded in Table 1 for each condition. Recall with topic and vehicle prompts was nearly perfect. This is not surprising since a topic or vehicle prompt supplies half of the sentence that must be recalled. However, it does indicate that nearly all of the sentences are available to subjects for later recall. Thus, these recall scores suggest an upper limit on how well recall might be prompted under the best conditions.

TABLE 1: Mean proportion of sentences recalled: Experiment I.

Acquisition list	Prompts Topics	Vehicles	Grounds A	Grounds B
A	.86	1.00	.70	.22
B	.86	.94	.26	.73

The results for ground prompts showed a strong interaction between Lists (A and B) and Grounds (A and B); this was verified in an analysis of variance for those four conditions [Lists x Grounds, $F(1,60) = 146.8$, $p < .001$]. There was no main effect for either Lists [$F(1,60) = 0.85$] or Grounds [$F(1,60) = 0.05$], suggesting that the lists were evenly balanced with respect to ease of recall and effectiveness of their related grounds in prompting recall. The source of the interaction is clear in Table 1: the grounds were effective as prompts only when subjects had heard the relevant acquisition sentence. In the case of Grounds A, recall of List A was far superior to recall of List B [$F(1,60) = 62.6$, $p < .001$]. With Grounds B, recall of List B was far superior to recall of List A [$F(1,60) = 85.0$, $p < .001$]. Similarly, from the standpoint of each acquisition list, relevant ground prompts were far more effective than grounds that were true of the topic but irrelevant to the sentence heard [$F(1,60) = 76.2$, $p < .001$ for List A; $F(1,60) = 70.6$, $p < .001$, for List B]. Overall, relevant grounds enabled subjects to approach perfect recall; recall was not far below the levels found for topic and vehicle prompts.

In this analysis, the matched pairs of acquisition sentences provided an internal control on the effectiveness of the grounds as prompts. Analysis of the variability associated with *subjects* showed that subjects performed best when the set of grounds was relevant to the set of acquisition sentences. Because of the crossed design of lists and prompts, the ineffectiveness of a set of irrelevant grounds could not be due to some flaw intrinsic to the prompts themselves (whether accidentally or by design): the same prompts were very effective when acquisition conditions were favorable. It is important to know for how many acquisition sentences and how many prompts this was true. If the effect was contributed by only a fraction of prompts that worked exceptionally well in relevant list conditions, then the results for subjects would be of far less interest.

To make an analysis of the behavior of *prompts*, we derived new scores from the original data by summing the number of subjects correctly recalling a sentence in each condition. The initial head count was impressive: 26 of the 28 acquisition sentences were better recalled with the relevant prompt than with the irrelevant prompt, and *all* of the 28 grounds were more effective in prompting recall of the *relevant* acquisition sentence.

To make a stronger test of these differences, we performed an analysis of variance for the behavior of prompts analogous to that performed above for the behavior of subjects. (We chose to study the variance associated with prompts, rather than acquisition sentences, since prompts were likely to show more variability and could be considered a repeated measure across list conditions, in each case providing a more sensitive test.) The mean proportions of subjects correctly recalling a sentence are equivalent to those in Table 1. Scores for the topic and vehicle prompts showed low variance, which verifies our earlier conclusion that *all* of the sentences are available for recall under optimal conditions. In the ground prompt conditions, there was again no main effect for either Lists [$F(1,26) = 0.67$] or Grounds [$F(1,26) = 0.03$], but there was a strong interaction between them [$F(1,26) = 115.7$, $p < .001$]. The source of the interaction was clear: prompts performed best when subjects had heard the relevant metaphors. All within-level cell mean contrasts in the Lists X Grounds matrix were significant. For each list, relevant prompts were superior [$F(1,26) = 36.3$, $p < .001$, List A; $F(1,26) = 33.6$, $p < .001$, List B]. For each prompt set, more subjects recalled sentences in the relevant list [$F(1,26) = 9.02$, $p < .01$, Grounds A; $F(1,26) = 12.2$, $p < .01$, Grounds B]. Thus, we can reject the hypothesis that the high recall in relevant prompting conditions was attributable to only a subset of prompts that (fortuitously or not) produced high recall. The results were general for each set of prompts. (We might add that the distribution of scores for each set showed no bimodality.)

A few prompts produced high recall of the related *irrelevant* sentence. For example, out of 16 subjects in the List A/Grounds B condition, 9 correctly recalled the *skyscraper-honeycomb* sentence when given the irrelevant ground, *are very tall compared to surrounding things*. This is apparently a case where the ground is so criterial a property of the topic that it is likely to remain invariant and salient no matter what the context of interpretation (see Lakoff, 1972). However, as the above analysis makes abundantly clear, such cases were exceptions to an otherwise consistent pattern: topics interpreted in one context tended to be inaccessible from other contexts.

Discussion

The results demonstrate that an abstract statement of the implicit ground of a metaphor is sufficient to remind a person of the metaphor at some later time. These abstractly related grounds were nearly as effective in prompting recall as the topics and vehicles explicitly mentioned in the sentences. The results are consistent with the hypothesis that subjects infer a resemblance during their initial encounter with a metaphoric sentence and that the resemblance is integral to what is stored as a memory of that experience. The interaction between lists and grounds further suggests that the semantic role of the topic is highly specific to the context supplied by the vehicle.

Before accepting these conclusions, however, there are other interpretations that must be considered. These are some of the alternatives, including one to which this study was directly addressed.

(i) Topic-property recognition. The vehicle does not interact in any way with the topic. The relevant ground is a (more or less salient) property of the topic. It prompts recall because the same property was activated during acquisition and formed part of the record of the event, or because in scanning a record of topics plus vehicles during recall, the system notes a match between the ground and the predicates or features already associated with the topic. According to this view, the vehicle is carried along as baggage, like the second term in a paired associate. A ground, therefore, should be as effective in prompting recall of "irrelevant" sentences (with the same topic) as it is in prompting recall of the relevant sentence. The experiment just reported indicates that this extreme position is untenable: the particular vehicle to which a topic is paired makes a substantial difference. Experiment II explores a more sophisticated version of this model.

(ii) Vehicle-property recognition. The relevant ground is more likely to be a salient property of the vehicle than of the topic. It prompts recall because the same property was activated by the vehicle during acquisition and formed part of the record of the event, or because in scanning through a record of topics plus vehicles during recall, the system notes a match between the ground and the predicates or features already associated with the vehicle. In this case, properties of the vehicle are seen as central to recall. The topic is carried along as baggage, and our understanding of it need in no way be transformed or enhanced. The particular topic to which a vehicle is paired should make little difference in the effectiveness of the relevant ground as a prompt for recall. This possibility will be tested in Experiment III, along with the possibility that the topic's or the vehicle's properties may provide the path for recall.

(iii) Topic or vehicle generation-recognition. Independent of any experience with the sentences, the likelihood is high that the relevant ground will make people think of the topic or the vehicle. It is sufficient to hypothesize that listeners make a kind of paired-associate record of the topic-plus-vehicle inputs, that the grounds lead them to generate many possible topics and vehicles at a later time, that they search their input list for a sentence that contained the generated term, and that they then output any sentence where they recognize a match. According to this view, one need not assume that properties of either the topic or the vehicle are activated at acquisition or compared at recall. It is an extreme form of the proposition that the topic and vehicle do not interact in any significant way. We will test this possibility in Experiment IV.

EXPERIMENT II

The results of the first experiment suggest that relevant acquisition experience facilitates the effectiveness of the grounds as prompts, and that irrelevant or conflicting experience interferes with their effectiveness. Therefore, one might propose a more dynamic version of the topic-property recognition model in which properties of the topic are primed or weighted differently in the presence of different vehicles. This involves relaxing the rather extreme constraint that the vehicle not interact with the topic, but preserves the assumption that pre-existing predicates or features of the topic are the basis for interpretation and recall. Prompted recall with relevant grounds would presumably be effective because those properties of the topics were specially primed, tagged, or weighted during acquisition. This change might coincide with a reduced weighting being given to other properties (including the irrelevant ground) and, in any case, it would presumably affect the recognition of other properties during recall.

This model can become very alluring, so we must keep its potential failings clearly in mind. The "property" under discussion may not be part of a person's knowledge before hearing a metaphor, and, even if it is familiar, it may have to be rediscovered with a nuance unique to that context. Metaphor not only brings us to see the unfamiliar, but to see the familiar in new ways.[3] The process of comprehension may involve more than activating a relatively stable network in a novel way or priming an unusual subset of features. It may involve a restructuring of the topic domain. Such a novel structuring would allow one to apprehend certain relations with ease, while other possible relations would be unavailable because apprehending them presupposes a different structuring. For example, the two metaphors about tree trunks invite us to structure our conception of tree trunks in entirely different ways. In contrast to the straws metaphor, the pillars metaphor leads us to conceive of trees as solid columns (rather than hollow tubes), to conceive of a forest of trunks (rather than an individual trunk), and to conceive of their function as holding up a solid mass of leaves and branches (rather than as transporting liquid to more individuated leaves and branches). We are not dealing with the same tree trunks in the two sentences, even though the isolated lexical items are identical.

In this experiment we studied the effect of the metaphoric vehicle by comparing subjects' interpretations of a topic with and without a vehicle.

---

[3] W. J. J. Gordon (1961), in his application of metaphoric thinking to problem solving, describes these functions epigrammatically as "making the strange familiar" and "making the familiar strange."

Subjects' recall of a list of isolated topics was prompted with the two sets of grounds, providing a measure of "comprehension" when no vehicles had affected interpretation of the topics. The overall design crossed three acquisition lists (A, B, and Topics Only) with two ground sets (A and B). Thus, for each set of grounds the following predictions could be tested.

(1) A ground should be less effective in prompting an isolated topic than in prompting a full metaphor with the relevant vehicle. This prediction would follow from any model that proposes interaction of the vehicle with the topic.

(2) A ground may be more effective in prompting an isolated topic than in prompting a full metaphor with an irrelevant vehicle. This prediction would follow if there is a greater likelihood that subjects will hit upon the "correct" context or properties while thinking about the isolated topic, compared to the topic in a conflicting context.

Method

A third acquisition list, List Topics Only, was recorded according to the same procedures used in recording Lists A and B. The list contained the topic noun phrases from the metaphors in the full-sentence lists, in the same order of appearance. Each topic was spoken twice, followed by a 5-sec pause. Two sets of prompt booklets, Grounds A and B, were identical in design to those used in Experiment I.

Subjects were 60 undergraduates in an introductory psychology course at the University of Minnesota. They received extra credit for their participation. Subjects were randomly assigned to one of three acquisition conditions: List A, B, or Topics Only. Ten subjects in each condition received Grounds A, while ten received Grounds B.

The procedure was the same as before. The experimenter read the acquisition instructions, and the subjects then listened to one of the three lists. In the List Topics Only condition the instructions were modified slightly; subjects were told they would hear a "short series of words and phrases" and were asked to "think about what each word or phrase is describing." Recall instructions were the same as before, except that subjects were asked to write down just the "topic" or "subject" of the original sentence (Lists A and B), or the "word or phrase" from the original list (List Topics Only). Thus, the recall tasks of all three groups were equalized to the extent that all subjects were to recall a phrase of equal length, and all responses could be scored according to the same criteria. Following the recall instructions, the experimenter distributed the prompt booklets and paced recall at 25 sec per prompt (compared to 40 sec in Experiment I, where both topic and vehicle were to be recalled).

Results

Topics were scored correct according to the same criteria used for accepting topics in full-sentence recall, that is, the response had to contain the central noun from the original topic noun phrase or a close synonym.

The mean proportion of topics correctly recalled by subjects is recorded in Table 2 for each condition. The pattern of recall in the two full-sentence list conditions replicates the pattern found in Experiment I. The level of recall for each group is also essentially the same as in the earlier experiment. Thus, it makes little difference to subjects whether they are asked to recall just the topic or the topic plus vehicle. If they can recall the topic, they will also be able to recall the vehicle with which its interpretation was (we presume) intimately connected.

TABLE 2: Mean proportion of topics recalled: Experiment II.

Acquisition list	Prompts Grounds A	 Grounds B
A	.69	.29
B	.21	.64
Topics Only	.41	.44

An analysis of variance for the six treatment groups in Table 2 showed no main effect of either Lists [$F(2,54) = 1.45$] or Grounds [$F(1,54) = 0.50$], but there was a large interaction between the two factors [$F(2,54) = 51.7$, $p < .001$]. One source of this interaction is familiar from Experiment I: relevant pairings of prompts with full-sentence lists produced high recall; irrelevant pairings produced low recall. For each acquisition list, relevant grounds were much more effective as prompts [$F(1,54) = 47.3$, $p < .001$, List A; $F(1,54) = 56.1$, $p < .001$, List B]. For each prompt set, relevant acquisition experience was far superior in facilitating topic recall [$F(1,54) = 67.7$, $p < .001$, Grounds A; $F(1,54) = 37.7$, $p < .001$, Grounds B].

A second source of the interaction is clear in a comparison of the recall for full-sentence lists and for the topics-only list. For each set of grounds, recall of the isolated topics was intermediate between recall of the same topics in the context of relevant vehicles and recall of the topics in the context of irrelevant vehicles. With Grounds A, recall for List A was superior to recall for List Topics Only [$F(1,54) = 22.9$, $p < .001$), which in turn was superior to recall for List B [$F(1,54) = 11.8$, $p < .01$). With Grounds B, subjects more successfully recalled the topics of List B than the same topics in List Topics Only [$F(1,54) = 11.8$, $p < .01$], which in turn were better recalled than the same topics in List A [$F(1,54) = 7.30$, $p < .01$].

Again, the generality of these findings needs to be verified by an analysis of the behavior of individual prompts. We need to be sure that the results are not the fortuitous contribution of a small subset of the metaphoric sentences and their grounds. An analysis of variance (taking prompts as a repeated measure across lists) showed no main effect for either Lists [$F(2,52) = 0.94$] or Grounds [$F(1,26) = 0.12$], but there was a large interaction between the two factors [$F(2,52) = 33.3$, $p < .001$].

The results for the full-sentence list conditions verified those found in the analysis of prompts in Experiment I. For each list, relevant grounds enabled more subjects to recall the appropriate topic [$F(1,26) = 11.7$, $p < .01$, List A; $F(1,26) = 13.9$, $p < .001$, List B]. For each set of grounds, relevant acquisition experience was superior in facilitating correct recall [$F(1,26) = 11.9$, $p < .01$, Grounds A; $F(1,26) = 6.60$, $p < .025$, Grounds B].

A second source of the overall interaction was found in contrasts between full-sentence and topics-only lists. However, the results of these contrasts were not as clear-cut as they were in the analysis by subjects. With Grounds A, the intermediacy of isolated topics between relevant and irrelevant metaphors was significant, but not strongly so. Topics of metaphors in List A were definitely recalled by more subjects than the same topics in List Topics Only [$F(1,26) = 9.84$, $p < .01$], but these in turn were only somewhat better recalled than the same topics in the metaphors of List B [$F(1,26) = 4.30$, $p < .05$). With Grounds B, for which within-group variances were especially high, the intermediacy of isolated topics was even less sharply defined: List B was better recalled than List Topics Only [$F(1,26) = 4.30$, $p < .05$], but List Topics Only was only marginally better recalled than List A [$F(1,26) = 3.13$, $.05 < p < .10$].

Study of individual prompts verifies the inconsistency of their behavior and suggests the source of any observed intermediacy of topics-only lists between relevant and irrelevant lists. While 24 (or nearly all) of the 28 grounds produced better recall of topics from relevant metaphors than from irrelevant metaphors, the recall of isolated topics was intermediate between the two in only 13 of the total cases. (We considered intermediacy to be any case where the number of subjects recalling a topic met the following inequality across the three list conditions: relevant list > topics-only list > irrelevant list.) The scores for individual prompts in the topics-only list condition showed modal values of three or four subjects recalling the topic, but the scores were spread throughout the range from 0 to 10 subjects. Most of the extreme cases happened to be prompts in Grounds B, accounting for the high variance in that condition. High recall apparently occurred when the ground was a salient or criterial property of the topic. For example, 10 out of 10 subjects recalled the isolated topic _skyscrapers_ in response to the ground _are very tall compared to surrounding things_. (Recall that this was also an effective prompt for the "irrelevant" _skyscraper-honeycomb_ sentence.) Low recall of isolated topics occurred when a ground required a relatively novel context for interpreting the topic. In response to the ground _are tubes which conduct water to where it's needed_, no subjects recalled the isolated topic _tree trunks_. Similarly, no subjects recalled _billboards_ in response to the ground _tell you where to find businesses in the area_. Apparently, the likelihood was very low that they would think of the relevant context during their original contemplation of the topic, or the likelihood was low that they could see the ground's relevance to the topic even if they scanned over the topic during recall. The power of a vehicle to lead subjects to discover this relevance is apparent in the recall scores for these same grounds in relevant list conditions: 6 out of 10 subjects correctly reported the topic *tree trunks* (having earlier heard the *tree trunks-straws* sentence), and 8 out of 10 subjects correctly recalled *billboards* (having heard the *billboards-yellow pages* sentence). The intermediacy in the topic-only list conditions apparently represents a central tendency along a continuum of likelihood that the relevance of a property will be noted in a null context.

Discussion

The results in the full-sentence list conditions support the claim that the vehicle plays a critical role in the comprehension and recall of metaphoric topics. If all properties of a topic could be activated at acquisition or recall, then any of them should serve to remind subjects of the topic. This was clearly not the case. With few exceptions, a specific property was a successful prompt only if it was integral to comprehension of the full sentence. When it was *not* integral to comprehending the sentence, subjects were only occasionally able to see its relevance to the topic at a later time.

The results in the topics-only list condition support this conclusion. If all possible properties were activated to an equal degree whenever the topic appeared, there should be no difference between isolated topics and any full metaphor containing them. But there were consistent differences; a particular property tended to be a good prompt for a relevant metaphor, variably intermediate for the topic alone, and a poor prompt for an irrelevant metaphor. Moreover, there was little correlation between the perceived relevance of a property to an isolated topic and its perceived relevance to the topic *in context*, as measured by prompted recall in each case. Across the 28 grounds, the correlation between the number of subjects recalling the topic from the relevant-sentence list to the number recalling it from the topics-only list was only 0.23.

These results do not support a simple form of the topic-property recognition model. A more sophisticated form of the model would need to propose how the vehicle enhances the saliency of one or more of the topic's properties. Models written in the framework of semantic feature theory and semantic network theory typically propose a search for common features or common associations (including associated predicates). For example, Johnson et al. (see footnote 2) and Malgady and Johnson (1976) compare metaphors to compound association stimuli and argue that features shared by the two nouns are raised in saliency, compared to non-overlapping features. They report that rated "figure goodness" correlates with the degree of rated similarity between the two nouns and the number of (independently assessed) shared attributes. Sternberg (1977) proposes that judgments about the validity of four-term analogies are based on component processes that include scanning for feature matches. Similar accounts can be written in terms of overlapping activation of predicates in a semantic memory model. Kintsch (1972, 1974), for example, suggests that the meaning of a metaphor is based on common "lexical implications" associated with its underlying terms.

All of these approaches assume that the ground of a metaphor is the logical intersection of two pre-existing sets of semantic elements, and that a sufficient comprehension strategy is to search for these common elements. An all-too-easy inference from these models is that sentences linking highly similar things in familiar contexts are quintessential metaphors: Skyscrapers are the giraffes of a city, and even Flowers are the blooms of a garden. Clearly, such a similarity continuum provides no basis for distinguishing metaphoric language from literal language or tautology, let alone for characterizing aesthetic quality.

While the common-elements approach appears to handle the most transparent comparisons, it is inappropriate for most of the sentences in this study. Properties that were poor prompts of the isolated topics cannot reasonably be said to be low-frequency or low-saliency entries in a pre-existing set of the topic's properties. We only become aware of such properties when a particular vehicle invites us to do so. We can add these properties, post hoc, to our list, but we will never be able to specify exhaustively all of the resemblances that we may potentially discover. Many studies of metaphor and analogy beg this question by using small preselected sets of attributes and values, and by making their identity obvious to subjects from the outset (for example, Sternberg, 1977). In natural contexts of metaphor or analogy use, the crucial task of comprehension is to discover what properties are relevant. The vehicle certainly plays a role in determining what is "relevant," but these constraints cannot be modeled effectively by a weighted matching function that selects out pre-existing attributes of the topic. As an account for all of the metaphors studied here, it may prove more parsimonious to say that "priming" results from a distinctive structuring of the topic domain for each metaphoric context in which the topic terms appear.

EXPERIMENT III

To this point we have considered properties of the topic as the focal point for processes in recall. The simple topic-property recognition model received negligible support. The specific vehicle paired with a topic exerts considerable influence on the topic's interpretation and its accessibility to recall at a later time. In cases where the ground is not part of prior knowledge about the topic, the vehicle's role in defining sentence meaning is clearly central. This leads us to consider a second possible class of featural explanations for the high level of prompted recall in relevant list conditions: vehicle-property recognition. In many cases the relevant ground is a salient property of the vehicle (considered in isolation). The use of such a vehicle presumably makes the metaphor more comprehensible and more effective in attributing a property to the topic. For example, the ugly protrusiveness of warts and the tallness of giraffes are both salient properties. The relevant grounds may be effective prompts because they specify properties that are activated when hearing the vehicle at acquisition, or that are easily discovered during some scanning process at recall.

There are various forms this hypothesis could take. Linguists and rhetoricians have often asserted that metaphor involves a transfer of meaning from the vehicle to the topic. (The Greek ancestor of the term "metaphor" meant to transfer or carry over.) In recent attempts to accommodate feature theory to metaphoric language, semantic interpretation is described as a transfer of part of the feature specification of the vehicle to the topic, adding and altering values in the feature specification of the topic (Weinreich, 1966; Bickerton, 1969; Leech, 1969; Thomas, 1969). In linguistic terms, this usually constitutes a more-or-less temporary alteration in the dictionary entry for the topic.[4] A similar process could be proposed in the framework of semantic memory models: the transfer would consist of adding a new predicate to the current representation of the topic. Orthodox behaviorists and mediationists might argue that metaphor is simply a case of classical conditioning. By pairing the topic and vehicle in close temporal contiguity, the ground (which is a strong unconditioned meaning response to the vehicle stimulus) may be transferred to the topic stimulus (see Osgood, 1953; Mowrer, 1954).

For each of the strong forms this hypothesis can take, the same conclusion follows directly: prompting of recall should be equally effective no matter what topic a vehicle is paired to, since the vehicle's properties determine the meaning and are the focal point for processes in recall. For the sentences in Experiments I and II, the vehicles were chosen to make comprehensible assertions about the topics (we will call these "principled metaphors"). The vehicle-property hypothesis suggests that the specific pairings should make little difference. Therefore, for this experiment we randomized the pairings of topic and vehicle phrases to create a new set of metaphoric sentences ("arbitrary metaphors"). If the relation of the vehicle to a ground is all that determines recall, then recall of these new metaphors should be as high as recall for the original metaphors. Only "relevant" list-grounds pairings were used in this experiment, for comparison with relevant prompted recall conditions in Experiment I.

Method

Two acquisition lists of arbitrary metaphors (Lists A' and B') were prepared from the principled metaphors by randomly reassigning pairs of vehicles to different topics. For example:

Tree trunks are like dragons.
Tree trunks are like babies with pacifiers.

Cigarette fiends are warts on the landscape.
Cigarette fiends are the yellow pages of a highway.

---

[4] Note, however, that metaphoric interpretations vary widely in permanency. Some metaphors request only a short-term orientation to a topic, as in the comparison of tree trunks to straws. Others presuppose more permanent (and more global) modes of orienting to the environment; for example, a tree trunk may be viewed as the residence of a malevolent being or as the umbilical of the Great Earthmother in a myth of biological genesis (Keeler, 1961). The duration of a metaphoric interpretation is another aspect of metaphor use that cannot be accounted for in terms of a user-independent axiomatic semantics.

The order of topics in each list was the same as in the comparable lists of principled metaphors (Lists A and B). The singularity/plurality of the topic and verb was adjusted in some cases to correspond to that of the vehicle. With this minor exception, the new lists contained the same verbal material as the original lists; thus, the memory tasks (simply conceived) and the possible intralist confusions were comparable. The lists were recorded under the same conditions as before; the intonation contours and pace were kept as natural as possible. Each sentence was repeated twice and was followed by a 5-sec pause.

The prompt booklets were identical in design to those used before (Grounds A and B). Thus, the order of correct recall of vehicles (and the topics paired to them) was the same.

Subjects were 20 undergraduates enrolled in an introductory psychology course at the University of Minnesota. They received extra credit for their participation. Subjects were randomly assigned to one of two conditions: 10 subjects heard List A' and received Grounds A as prompts, and the other 10 heard List B' and received Grounds B.

The listening conditions and acquisition instructions were the same as before. The experimenter mentioned that some of the sentences would be a little bizarre and asked subjects to do their best to find sensible interpretations. Recall instructions were those used in Experiment I, that is, subjects were asked to recall the full sentence most related to each prompt, as well as they could remember it. They were paced at 40 sec per prompt.

Results

In scoring subjects' responses for the appearance of topics and vehicles, the same criteria were used as in previous experiments. In the initial scoring procedure, the sentence containing the vehicle originally related to the ground was judged to be the "correct" sentence to recall. Both the topic and the vehicle of this sentence had to be correctly recalled.

The mean proportion of arbitrary metaphors recalled per subject is recorded in the second column of Table 3 for each list condition; the results for principled metaphors in comparable conditions are included in the first column for comparison. The results were clear: when a vehicle appeared in a principled metaphor, relevant prompted recall of the sentence was substantially greater than when the same vehicle appeared in an arbitrary metaphor. This difference was significant for both sets of grounds [two-tailed $\underline{t}(24) = 4.04$, $\underline{p} < .001$, Grounds A; $\underline{t}(24) = 5.53$, $\underline{p} < .001$, Grounds B]. This rules out any simple hypothesis that ascribes relevant prompted recall solely to the relation between the ground and the vehicle.

TABLE 3: Mean proportion of sentences recalled: Experiment III.

		Arbitrary metaphors		
Prompts	Principled metaphors[a]	Vehicle sentence	Topic sentence	Topic or Vehicle sentence
Grounds A	.70	.40	.11	.51
Grounds B	.73	.34	.16	.50

[a] From Table 1, Experiment I.

In the previous experiments we considered a complementary hypothesis that ascribed recall solely to the relation between the ground and the topic. The present results allow another test of that hypothesis. Subject's responses were rescored, counting as "correct" any sentence that contained the *topic* originally related to each ground prompt. The mean proportion of sentences correctly recalled by subjects according to this criterion is recorded in the third column of Table 3 for each condition. A sizable fraction of the arbitrary metaphors correctly recalled by subjects resulted from a close relationship between topics and grounds. Even so, the fraction attributable to topics was substantially smaller than that attributable to vehicles. Topic-property recognition is even less successful than vehicle-property recognition as a predictor of the level of recall for principled metaphors.

We are now in a position to test a combined hypothesis: the recall of metaphors may involve prompting of either the topic or the vehicle (by means of an associated property that matches the ground), followed by recall of the other member of the pair. A comprehension process laying the groundwork for this recall process could be framed in terms of probabilities or saliencies. There may be a certain probability that an appearance of the topic will activate a relevant property, and an independent probability that the vehicle will activate the same property. There may be a certain resting saliency of the property in the topic domain and an independent saliency in the vehicle domain. The possible success of a combined hypothesis is suggested by results for some of the arbitrary metaphors. In the few cases where a topic sentence was frequently recalled, the ground tended to be a salient property of the topic; for example, 4 out of 10 subjects recalled the *skyscraper-branding iron* sentence in response to *are very tall compared to surrounding things*. In cases where a vehicle sentence was frequently recalled, the ground tended to be a salient property of the vehicle; 9 out of 10 subjects recalled the *cigarette fiends-warts* sentence in response to *are ugly protrusions on a surface*.

Whether the combined model is phrased in terms of prior probabilities or saliencies, the critical assumption is that the values associated with the topic and vehicle domains are independent. If probabilities related to the vehicle are zero, the model reduces to a topic-recognition model. If probabilities related to the topic are zero, we have a vehicle-recognition model, and it is irrelevant whether we choose to speak of "transfer" of properties to the topic. If both probabilities are nonzero, we have the model described at the end of Experiment II: the ground of a metaphor is the intersection of two independent property sets. The relation between the ground and the metaphor will be characterized by a *joint* probability in addition to the probabilities associated with the topic and vehicle alone. This model, in the language of saliencies, is best exemplified by the work of Johnson et al. (footnote 2) and Malgady and Johnson (1976).

The combined model asserts that the probability of recall of principled metaphors is the sum of the probabilities for prompting only topic recall, only vehicle recall, and both topic and vehicle recall. (This assumes that the probability is unity of getting from only the topic or the vehicle to the full sentence; the results of Experiment I indicate this is a reasonable assumption.) The recall data for arbitrary metaphors do not allow us to estimate these three probabilities directly, since we do not know how subjects divided their responses between the topic and vehicle sentences when both came to mind. However, we can estimate the *total* probability by summing the topic and vehicle sentences recalled by each subject and averaging the new set of scores. These estimates are recorded in the fourth column of Table 3. For each set of grounds, the mean for topic or vehicle sentence recall was significantly less than the mean for principled sentence recall [$t(24) = 2.33$, $p < .05$, Grounds A; $t(24) = 2.80$, $p < .01$, Grounds B].[5] In addition, at the level of individual prompts, there was *no* correlation between the frequencies of recall for arbitrary and principled metaphors ($r = 0.005$ for the 28 grounds). Thus, a combined model, assuming independently defined probabilities or saliencies for the topic and vehicle, is not adequate as a predictor for the recall of metaphoric sentences and, by implication, may not be adequate as an explanation for their comprehension.

Discussion

It is possible to accept this conclusion without negating the intuitions that motivated the models tested here. For example, the importance of salient aspects of the vehicle domain seems unquestionable. The vehicle exerts a tremendous influence on the accessibility of principled metaphors to recall, and it is clearly the more common pathway for recall of arbitrary topic-vehicle combinations. Thus, the comprehension of metaphor may involve a presupposition that the dominant source of constraints on meaning is the vehicle, and that the topic should be comparatively malleable to interpretation. Even if one argues for a mutual influence of topic and vehicle domains on each other, it seems clear that the degree of influence is asymmetrical.

---

[5] It should be noted that almost all of the sentences correctly recalled were either topic sentences or vehicle sentences. Thus, the lower total recall for the arbitrary metaphors cannot be attributed to the intrusion of incorrect responses. The number of intrusion errors in Experiment I was similarly small.

This again raises the question of independence and interaction. With the exception of the more extreme vehicle-property transfer theorists, almost everyone would agree that the topic and vehicle "interact" in a comprehender's interpretation of metaphor, in the loose sense that both affect the resulting meaning. There are two levels, however, at which the question of independence needs to be posed. At the more fundamental level, we must ask whether the topic and vehicle are "separable." This is a question about what hypothetical entities provide the most useful basis for an explanatory theory of the process of comprehending metaphor. If we assume the topic and vehicle to be separable, then we are assuming that they have associated properties, probabilities, saliencies, states, or processes that are independently defined. Having assumed distinct entities at this level, we can proceed to ask whether the two sets of entities interact in the hypothetical processes underlying comprehension. Most of the current linguistic and psychological approaches to semantic interpretation assume separability of the entities attributed to individual words: their features, concepts, predicates, meanings, associations. For example, Johnson et al. (see footnote 2) attribute distinct feature vectors to each term and then define the meaning of the full metaphor in terms of the union and intersection of these two feature vectors. They make a point of asserting that this is an "interactive" process, and, in a secondary sense, it is; but at the fundamental level their model assumes that the two terms function independently and additively. A comparable distinction would apply to semantic network accounts of metaphor; these models assume separate storage of information for each domain and define semantic interpretation in terms of new interconnections.

The assumption of separability is a natural one. We perceive words and objects as having separate identities, and it is natural to try to characterize these identities in isolation. Dictionaries serve useful functions, and it is tempting to assume that hypothetical dictionaries (lexicons or networks) will provide a sufficient base for hypothetical processes of comprehension. The crucial question for cognitive theory is whether words are functionally separable. In the pursuit of meaning, in response to sentences and longer discourse, the cognitive impacts of component words may be only partially separable.

The results for arbitrary metaphors provide a strong (though certainly not definitive) test of models assuming separability of words and a more-or-less additive process for their combination. To these models, all topic-vehicle combinations are fundamentally arbitrary. However, it is clear from the data that "arbitrary" pairings do not have the cognitive force of "principled" pairings (intuitively defined). Subjects' performance on arbitrary pairings did not provide adequate estimators for their performance on principled pairings. It is also worth noting that the frequency of recalling only a topic or a vehicle was substantially higher for arbitrary metaphors than for principled metaphors. Recalling the topic or vehicle of an arbitrary metaphor does not always allow recall of the other member of the pair; thus, the assumption made above that this probability is unity does not hold for arbitrary pairings. This suggests that subjects' representations of arbitrary pairings are less integrated; they have been forced to deal with many of the topics and vehicles as separate entities. One further symptom of this is the appearance of combinations in recall that were not heard during acquisition. In response to are very tall compared to surrounding things, one subject responded with a sentence combining two topics: Skyscrapers are billboards to a large city. Another subject recombined two pieces to produce the original principled metaphor: Skyscrapers are the giraffes of a city. In addition, four subjects recalled the related topic sentence (skyscrapers-branding irons), one recalled the vehicle sentence (matches in a forest-giraffe), one recalled only the topic (skyscrapers), and two recalled only the vehicle (giraffes).

To a language user, the "same term" is not the same term in each context of combination. The "same vehicle" need not have the same predicating potential in all contexts. A predicate that is an effective prompt in one topic context (principled metaphors) need not be effective in another topic context (arbitrary metaphors). Similarly, the "same topic" is not functionally the same when combined with different vehicles. The possible relevance of a predicate to a topic may be perceptible only if the topic has appeared in the context of a particular type of vehicle. As argued above, this kind of flexibility in a term's function is true of all language use and cannot be characterized by prescriptions in a lexicon. The crucial question for metaphor is not what constraints need to be relaxed, but what constraints need to be imposed to make metaphoric combinations interpretable. The topic and vehicle are not totally flexible; arbitrary combinations are not as easily integrated as principled combinations. The reason for this may be the receptiveness of the topic to the "structuring" suggested by the vehicle (assuming the vehicle plays the dominant role). We can easily transform a tree trunk into a straw or a pillar, but not so easily into a dragon or a baby with a pacifier. It is doubtful that a logic of topic-vehicle compatibilities can be successfully framed in terms of elemental semantic features or predicates. The process of comprehension involves a more global transformation of the topic domain. Compatibility with a vehicle depends on the susceptibility of the entire domain to the appropriate transformation, and each such transformation defines new "properties" for the topic. It is in this sense that the topic's semantic structure is not fundamentally separable from the vehicle.

These considerations lead us to suggest that the comprehension process results in a partial identification (or fusion) of the topic and vehicle domains. To some extent, the imagined tree trunk may become a straw and the skyscraper may become a giraffe extending its neck above the city skyline. This mode of comprehension may be more common and integral to adult language use than is currently recognized. It has typically been assumed that "identification" is uniquely characteristic of pathological, poetic, or primitive thought; for example, the "paleologic" thinking of schizophrenics (as defined by Arieti, 1974), "primary process" thinking (for example, Freud, 1950), poetic imagination (Richards, 1960; Hawkes, 1972), symbolic play in children (Piaget, 1962; Gombrich, 1968), and magical thinking. While healthy use of metaphor does not typically entail a total identification of the topic and vehicle, the assumption of full functional separation seems equally extreme. Productive use of metaphor in problem solving, scientific theory, poetry, and personal growth probably demands a partial fusion of the two domains.

EXPERIMENT IV

The models discussed in the previous experiments assume that particular properties are apprehended during the process of comprehension, and that they later determine the accessibility of the topics and vehicles. We now consider an alternative approach that resists postulating such properties as mediators and attributes recall to a "direct" relationship between the grounds and the relevant topics and vehicles. For example, the phrase are ugly protrusions on a surface might lead subjects to think of warts independent of any special acquisition experience involving inference, matching, pairing, or other postulated processes. Prompted recall could consist of generating possible terms (for example, warts) in response to the prompt, searching some record of the original sentences until a matching term is recognized, and then reporting the sentence containing it. This recall procedure is similar to the "generation-recognition" model tested by Tulving and Thomson (1973) in their analysis of prompted recall for word lists, and it has been suggested by Osgood[6] as a possible explanation for the data reported here. In its simplest form, the model treats a metaphor as an uninterpreted paired associate that is stored in an "episodic memory" (Tulving, 1972) for later recall. While this is not a satisfying explanation of what it means to understand a metaphor, it could be sufficient to account for our earlier data in relevant prompted recall conditions.

To test this possibility we need an estimate of how likely people are to think of the relevant topic or vehicle when they read a ground without any prior experience with the acquisition sentences. To make these estimates we devised the following sentence completion task.

Method

Two sets of mimeographed response booklets, Grounds A and B, were prepared. They contained the grounds for Lists A and B, respectively. A cover sheet informed subjects that their booklets contained some incomplete sentences. They were asked to complete each sentence by supplying a "subject," using either a single word or an extended phrase. They were asked to write down at least three possible subjects and to work quickly, recording their answers as soon as they came to mind. The following example was provided.

_______________are very colorful.

1. Flowers
2. Hawaiian shirts
3. Eccentric people

---

[6] Osgood, C. E. (November 28, 1973): personal communication.

The order of the phrases in each form was the same as in the prompt booklets used in earlier experiments.

Subjects were 64 undergraduates enrolled in introductory psychology courses at the University of Minnesota. They were randomly assigned to one of two groups, receiving Grounds A or B. Approximately half of each group received extra credit for their participation; the remainder completed the form as a class assignment. Subjects worked individually in a quiet experimental room or classroom.

Results

Responses to each ground were scored as "topics" or "vehicles" if the terms were identical to or closely synonymous with terms in the original topic and vehicle phrases of the relevant metaphor. For example, moles and pimples were also accepted for the vehicle warts; beehives was accepted for honeycombs; and IDS building (the skyscraper in Minneapolis) was accepted for skyscrapers. Separate tallies were made for topic and vehicle responses; only the first appropriate response of each type was recorded.

The mean proportion of topics and vehicles produced by subjects is recorded in Table 4 for each set of grounds. On the average, subjects were more likely to think of related vehicles than topics by a factor of about 2:1. This bias toward vehicle responses is similar to that observed in Experiment III and suggests a complementary hypothesis about why particular vehicles are chosen as metaphoric predicates: they are exemplary instances of particular relationships. When encountering a ground under free association conditions, subjects are more likely to think of the vehicle domain (where the relationship is familiar) than the topic domain (where its relevance may not be familiar).

TABLE 4: Mean proportion of topics and vehicles produced in sentence completion task: Experiment IV.

Set of grounds	Topics	Vehicles	Topics or vehicles
Grounds A	.05	.18	.22
Grounds B	.12	.17	.28

However, these domains are only two among many that are likely to come to mind. The question is whether they do so often enough to account for the level of relevant prompted recall in earlier studies. The third column in Table 4 records the mean proportion of topics or vehicles supplied by subjects for each set of grounds.[7] On the average, subjects thought of 25 percent of the topics or vehicles. If being reminded directly of the topic or vehicle were a prerequisite for recall of principled metaphors, then we could expect subjects to recall no more than 25 percent of the 14 sentences, even if we assume recall proceeds without error once a topic or vehicle is known. This estimate falls far short of the level of relevant prompted recall observed in Experiment I, where subjects were able to recall about 72 percent of the sentences [$\underline{t}(94) = 14.4$, $p < .001$, for the two sets of grounds combined].

Not surprisingly, this finding is repeated in an analysis of grounds. For each ground in the sentence completion task, one can score how many subjects (out of 32) responded with the related topic or vehicle. The mean proportions of subjects are equivalent to the means in Table 4 and lead to a complementary conclusion: the probability that a topic or vehicle will be produced in response to a ground is substantially higher when subjects have heard the relevant acquisition sentence. This suggests a more sophisticated form of the generation-recognition hypothesis. The acquisition sentence may prime the topic and vehicle, making it more likely that they will be evoked during recall as implicit responses to the ground. If this priming is exerted equally by all topics and vehicles in the acquisition list, then the sentence completion data should enable one to predict the relative probability of prompted recall for individual grounds. For example, grounds that frequently evoke topic or vehicle responses in the sentence completion task should also produce high levels of correct recall in the prompted recall task. In other words, there should be a strong correlation between a ground's behavior in the two tasks.

A test of this hypothesis is facilitated by the substantial variability among grounds in each task. Experiment I measured the probability that each ground would produce correct recall of the full relevant sentence. We may take these as observed probabilities and test the power of an associative model to predict their configuration. Rough estimates of associative probabilities may be obtained from the proportion of subjects producing the topic or vehicle in response to each ground. These estimates assume that recall proceeds errorlessly if either the topic or the vehicle is implicitly generated.

Observed and estimated probabilities showed little systematic relationship. For the 28 grounds, the coefficient of correlation between these estimated probabilities and the observed probabilities was only 0.17. This comparison assumes that priming is a linear function of extra-experimental associative probability. If priming is assumed to preserve linearity of the logarithm of probability measures, the correlation remains low and nonsignificant ($r = 0.27$). Thus, the associative model outlined above cannot successfully predict either the overall level or the specific configuration of relevant prompted recall.

---

[7] Inclusive or. Note that each figure is smaller than the sum of probabilities for topic and vehicle responses, since subjects occasionally responded with both. It is worth noting that the probabilities of responding with the topic and the vehicle are independent. The estimated probability of topic/vehicle co-occurrence would be (0.054) (0.176) = 0.0095 for Grounds A and (0.123) (0.174) = 0.021 for Grounds B. The mean observed probabilities of co-occurrence were not significantly greater than these estimates; the observed values were 0.0089 for Grounds A [$\underline{t}(31) = 0.11$], and 0.016 for Grounds B [$\underline{t}(31) = 0.93$]. This suggests there was little or not pre-existing "associative strength" between the topics and vehicles of the original metaphors.

More sophisticated probability estimates would acknowledge that recall may not proceed errorlessly if only the topic or the vehicle is generated. In Experiment I, there was some variability in the effectiveness of topic and vehicle prompts, and topics were slightly less effective overall than vehicles. A more accurate predicted probability for each ground could be obtained using the following equation:

$$\hat{p} = p(T).p(S/T) + p(V).p(S/V) + p(TV).p(S/TV),$$

where $p(T)$ is the probability of responding associatively with *only* the topic, $p(S/T)$ is the probability of producing the full sentence given the topic, $p(V)$ is the probability of responding associatively with *only* the vehicle, $p(S/V)$ is the probability of producing the full sentence given the vehicle, $p(TV)$ is the probability of responding associatively with *both* the topic and the vehicle, and $p(S/TV)$ is the probability of producing the full sentence given both the topic and the vehicle. Estimates of $p(T)$, $p(V)$, and $p(TV)$ for each ground were obtained in this experiment (using a measurement scale of 32 subjects).[8] Estimates of $p(S/T)$ and $p(S/V)$ for each ground were obtained in Experiment I (using a much coarser scale of eight subjects). $p(S/TV)$ may be assumed to be 1.00. Across the 28 grounds, the correlation of $\hat{p}$ with the observed probability of relevant prompted recall was only 0.18. Thus, the more careful estimation procedure does not alter the original conclusion: the generation-recognition model cannot predict the configuration of prompted recall.

It is worth noting that in a few cases the original vehicle was a frequent response to the ground in the sentence completion task; for example, *warts*, *pimples*, and the like were common responses to *are ugly protrusions on a surface* ($\hat{p} = 0.68$), and *yellow pages* was a common response to *tell you where to find businesses in the area* ($\hat{p} = 0.50$). In one case the original *topic* was a common response to the ground: *skyscrapers* and *IDS building* were frequent responses to *are very tall compared to surrounding things* ($\hat{p} = 0.69$). In these exceptional cases, the original vehicles or topics happened to be the most salient instances of the relationship specified abstractly by the ground, and the estimated recall probabilities approached the observed values. In general, however, responses to the grounds showed little correspondence (in either absolute or relative frequency) to the topics and vehicles produced in relevant prompted recall.

---

[8] Note that these estimates require rescoring the original data. Earlier we scored the number of subjects producing a topic or a vehicle (irrelevant of whether the other term co-occurred in individual subjects' responses). $p(T)$ requires scoring responses which include *only* the topic, $p(V)$ involves responses which include *only* the vehicle, and $p(TV)$ is the probability of co-occurrence. This breaks down the earlier "rough" probability estimate (total topics or vehicles) into three components.

Discussion

The results of this experiment demonstrate that the hypothesis of pre-existing associations between grounds and topics/vehicles provides little explanatory power. Neither the overall level nor the specific configuration of recall can be accurately estimated from the strengths of such associations. At the very least, this confirms our intuition that recall of a metaphoric sentence cannot be ascribed to a direct prompting of component terms, but involves some kind of match between relationships experienced at the invitation of those terms and the relationships specified by the ground. The product of comprehension must be more than a novel paired associate, more than a new "link" between the two terms or two classes of objects.

Tulving and Thomson's (1973) discussion of paired-associate stimuli applies to some extent to the conjunctions of noun phrases in metaphors: while the "nominal memory unit" is no more than a conjunction of terms, the "functional memory unit" can be a much more elaborated cognitive product. It is the functional unit that governs accessibility of the terms to later recall. In the case of metaphor, the functional unit can be an elaborated event or structure in which the terms' referents are only local components. The relationship between the ground and this elaborated structure exerts a greater influence on recall than any pre-existing relationship between the ground and the particular components mentioned in the sentence.

The logic of this experiment was complementary to that of previous experiments, but led to similar conclusions. Models tested in the earlier experiments assumed the prior existence of stored predicates or features that would be activated during comprehension. These properties were assumed to provide a sufficient set of constructs for characterizing the resulting meaning and the possible entry points for recall. With few exceptions, the distinctive relationships between metaphors and grounds could not be explained satisfactorily by these models. In contrast to these models, which assumed strong "forward associations" between sentence terms and properties, the generation-recognition model tested in this experiment assumed strong "backward associations." Again, the distinctive relationships between metaphors and grounds could not be accurately predicted. The relationship created by metaphor has nothing necessarily to do with familiar ways of structuring knowledge.

To the extent that the strengths of the postulated forward and backward associations show some correspondence, this experiment could be viewed as a replication of Experiment III. The convergence of the two experiments is suggested by the similar distributions of topic and vehicle responses (see Tables 3 and 4) and the similar interaction with sets of grounds (A and B) in each case. It is possible that arbitrary metaphors more closely fit the assumptions of the generation-recognition model than principled metaphors.

There were suggestions that the terms in the arbitrary metaphors often did not interact in the specification of meaning, that the terms were more available for recall as isolated and interchangeable units, and that they were more likely to be interpreted in terms of normative properties. However, the failure to find a correlation between recall of principled and arbitrary combinations could have been due simply to the fact that different metaphoric combinations specify different grounds. The interpretation of arbitrary metaphors could be as novel and interactive as that of principled metaphors. If so, the sentence completion data should be no better as predictors for the arbitrary metaphors than they were for the principled metaphors. On the other hand, if the behavior of arbitrary metaphors is much more a consequence of normative properties of their component terms, then the estimated probabilities based on the sentence completion data may have greater predictive power.

Results suggest that prior associative connections play a much greater role in the recall of arbitrary metaphors. Across the 28 grounds, there was a significant correlation between frequency of topic responses (Experiment IV) and frequency of recall of topic sentences (Experiment III), $r = 0.42$, $p < .05$. The correlation between frequency of vehicle responses and frequency of recall of vehicle sentences was even stronger, $r = 0.55$, $p < .01$. Finally, we can consider the combined recall for arbitrary topic and vehicle sentences. The observed frequency of recall and the total estimated probability ($\hat{p}$) of recall showed a significant correlation, $r = 0.48$, $p < .02$. Thus, the results for arbitrary metaphors and free association to grounds are significantly correlated with each other, but neither set of results is closely related to the behavior of principled metaphors. Prior associative connections (whether forward or backward or both) apparently play little role in the comprehension and recall of nonarbitrary metaphoric sentences.

GENERAL DISCUSSION

These experiments gave no indication that metaphoric comprehension is a specialized skill in which only certain people excel, or that metaphoric sentences are especially difficult to comprehend. Our listeners showed no bimodality in recall performance, and their average level of recall in relevant prompting conditions was very high. If metaphoric comprehension is a skill in deviance, it is a normal one.

We have taken the high level of relevant prompted recall as evidence that listeners discerned an abstract resemblance between the topic and vehicle domains. A paraphrase of the ground was highly effective as a prompt, even though the resemblance was not explicit in the original sentence, and the prompt contained no content words from the sentence. The results of Experiment IV indicated the necessity of postulating this implicit resemblance as a central component of comprehension and a mediator for recall; direct associative connections between the prompts and acquisition sentences could not predict the configuration of prompted recall performance. Subjects' paraphrases in recall provided further evidence for the presence of these grounds in their interpretations. They occasionally added to or modified the original terms, making it clear that they had inferred the appropriate resemblance: *Tree trunks are like straws* that give drink to *the leaves*; *Smokers are like* fire-breathing dragons.

These results have raised several issues concerning the structure of metaphoric resemblances, the process of comprehension, and the process of recall. In each case, we would like to sketch an alternative to attributive models that seems more consonant with our empirical findings and more fruitful as a vehicle for future theory and research. We hope this bold sketch will open avenues of investigation by which all models may become better articulated.

The Structure of Resemblance

In our discussion of the individual experiments, we considered various means of characterizing the grounds of metaphoric sentences. For both empirical and theoretical reasons, we have chosen to characterize metaphoric grounds in terms of abstract relations, rather than attributive features. We found negligible support for recall models that postulated the recognition of pre-existing attributes associated with topics, the priming or weighting of such attributes during acquisition, or the transfer of salient attributes associated with vehicles. While other models of this class could certainly be designed, we found no reason to believe that these were steps in the right direction.

A central question in this discussion is how the ground is related to the nominal terms of a metaphoric sentence. (We will limit ourselves here to sentences of the form "A" is (like) B," where A and B are both noun phrases.) Attributive models characterize the nominal terms by a list or array of features, and they characterize the ground by some weighted function of these features. These models are not well suited for characterizing grounds when the resemblance is not between the two terms (objects) per se, but between events or relationships in which each participates. Therefore, we prefer to describe metaphoric resemblances as relations between topic and vehicle domains (or schemata). Each domain is an abstract relationship among several entities; only a subset of these entities appears explicitly as nominal terms in the sentence. Thus, it is not strictly appropriate to identify the topic or vehicle of a metaphor with specific terms appearing in the sentence. In the tree trunks-straws sentence, for example, the topic term is tree trunks, but the topic domain is a type of transformation (fluid transport) exerted over certain structures (tree trunk, leaves and branches, water, roots, earth, etc.). A comparable description is also necessary for the vehicle domain, which is only partially specified by the terms straw amd thirsty. The ground combines the transformational invariants (for example, suction, fluid flow) and structural invariants (for example, vertical cylindrical space) that are common to each domain.

A semantic characterization of nominal terms must be made in a way that facilitates achieving a topic domain, vehicle domain, and transformational/structural resemblances as the "product" of comprehension. Simply activating a set of normative, context-free, structural descriptors is not enough (inanimate, cylindrical, plastic, hollow, 6-10 in. long, etc.). It seems preferable to suppose that a nominal term can activate a system of abstract structural and transformational invariants (that is, a domain or schema). These invariants will conjointly specify constraints on the relationships that the nominal term can participate in. The semantic characterization may also include particular instantiations of these abstract constraints within normative contexts. For example, the term straw could activate the following system of abstract constraints: a structure of relatively rigid nonporous material, of a hollow cylindrical shape, with a small diameter relative to its length. This structural specification is compatible with the accompanying transformational specification of event(s) within which the structure participates: the vertical cylindrical space channels fluid flow from a receptacle to a destination against gravity; the goal of the fluid transport is to alleviate thirst; the force for the flow is suction. In its normative contextual instantiation, the structure is paper or plastic, the receptacle is a bottle or cup, the destination is a person (the thirsty agent), and the source of suction is the person's mouth and lungs.

The Process of Comprehension

Given this speculative characterization of the knowledge activated by nominal terms, we now consider the role played by these terms in the process of comprehension. We have noted several indications that the vehicle plays the major role in guiding the comprehender toward a resemblance. Schemata in the vehicle domain tend to be the predominant source of constraints by which the topic domain is interpreted. In the tree trunks-straws sentence, for example, the comprehender is invited to apply the straw schemata to the tree trunks domain, that is, to create similar relational systems among appropriate entities in the new domain. In this creative process of schematization, the comprehender will seek to instantiate both the transformational and structural aspects of the vehicle domain: the trunk as the vertical cylindrical space, the leaves and branches as the thirsty agents and source of suction, the earth as the receptacle, ground water as the fluid, the transport of water as the fluid flow, etc. This process will lead to a growth in knowledge when the topic domain is successfully organized by schemata that are unfamiliar or unconventional in that context. The activation of knowledge by topic and vehicle terms is apparently asymmetric: the topic terms activate a comparatively unconstrained system of potential relationships, while the vehicle terms activate specific schemata that are more tightly constrained. Rather than relaxing normative constraints on the topic, the comprehender seeks to impose specific constraints from the vehicle domain, so that the topic term (object) participates in a specific type of event or relationship characteristic of the vehicle. This model of the comprehension process predicts a marked "specificity of encoding" for topic terms, a prediction that is consonant both with our prompted recall data and with the recall of nonmetaphoric materials (for example, Tulving and Thomson, 1973; Bransford and McCarrell, 1974).

At this point we have been able to provide only a rough framework for a model of the comprehension process. More explicit formulations will become possible as solutions are found to several remaining puzzles. One puzzle is how the terms in a metaphoric sentence activate the vehicle domain. The single nominal term straws, for example, clearly underspecifies all of the structures and events in the elaborated vehicle domain. One factor that shapes the resulting domain is the "familiarity" or "salience" of certain events or relationships in which the object can participate (though this does little more than label the phenomenon). The results for both the arbitrary metaphor and the sentence completion tasks provided circumstantial evidence that vehicles are more likely than topics to be exemplary instances of the grounds, and, conversely, that the grounds are more likely to be salient schemata for vehicles than for topics. Another factor is the use of contextualizers to constrain the comprehender's search for the intended schema. For example, finding the appropriate schema for straws is aided by extending the predicate phrase to are straws for thirsty X. Also of great importance is contextualization of the topic. Topic terms often appear mixed into the predicate, as in thirsty leaves and branches (tree trunks), giraffes of a city (skyscrapers), and warts on the landscape (billboards). These phrases aid in delimiting the appropriate schema and lead listeners to supply comparable entities in the vehicle domain. This was evident in paraphrases like the following (where even the ordering of topic and vehicle was reversed): Giraffes are skyscrapers of the jungle; Giraffes with other animals are like the skyscrapers in the city. Thus, it is not sufficient to argue that the topic is "passively" schematized by salient properties of a vehicle domain: the topic and vehicle terms interact in specifying the ground (see Black, 1962; Verbrugge, 1977).

A second puzzle for future research is to identify the constraints that govern successful schematization. The topic domain does not accept all transformations with equal ease. It is easier to schematize tree trunks as straws than as babies with pacifiers. There must be compatibility constraints operating between the topic and vehicle that govern what relations from the vehicle domain can be extended successfully or easily. These compatibility constraints, defined over abstract relations, may play a major role in judgments about metaphoric force and quality. Attributive conceptual theory has sought to define these constraints in terms of weighted conventional attributes and typically defines grounds as novel attributes transferred to the topic. But simply attaching new labels to a topic term does not provide a basis for determining when the process proceeds easily or successfully. The attributes represented in an attributive concept are properties that an object manifests in a heterogeneous set of conventional events or relationships. We are doubtful that a metric defined over such attribute lists can predict the ease of interpreting the topic in an unconventional event or relationship. Such a prediction may be possible only for transparent and uninformative metaphors (such as the skyscraper-giraffe sentence). We suspect that it will prove easier to define constraints on metaphoric transformations if structural concepts are defined from the outset by potential transformations under which they remain invariant. As we noted above, this may allow theoretical development of a single type of comprehension process that generates interpretations for both metaphoric and literal sentences.

A third major puzzle is how to characterize the topic domain so that it has sufficient functional plasticity to allow for novel schematization, yet is sufficiently constrained that various vehicle domains are differentially compatible with it. Models based on normative associations do not have sufficient plasticity to explain how the topic domain can be schematized in radically different ways in the context of different vehicles. Associative network models, semantic feature theories, and models of attributive conceptual knowledge all seek to interpret novel sentences by reference to fixed connections established over long experience. Such systems grow only by accretion; radical transformations, contingent on specific contexts, are not normally envisioned or easily modeled. Our results suggest that the topic domain is highly malleable as a function of the vehicle context; a topic is not "recognized" during recall unless the ground specifies the relationship by which it was originally schematized. To accommodate metaphoric growth in a general theory of comprehension, we need to characterize semantic structures by systems of organization that allow for greater functional plasticity than is possible in heterogenous networks and hierarchies. (See Turvey, Shaw and Mace, in press, for discussion of an analogous problem.)

The Process of Recall

If metaphoric grounds are characterized as abstract relations, their effectiveness as prompts poses a challenge for current models of the recall process. Experimental studies of word and sentence memory have emphasized the identities of the terms encountered during acquisition. It is assumed that these are central to the cognitive representation of the event and serve as the focus for organizational processes and recall. Verification probes, recognition foils, and recall prompts usually contain terms that appeared in the original event or terms "associated" with the acquisition terms in earlier experience. Our results, like those of Tulving and Thomson (1973), suggest that acquisition terms do not have a stable specific identity or set of associations in different contexts of interpretation. A prompting event may "identify" the related acquisition event by means of an abstract transformational resemblance. A relation of nominal or associative identity is not necessary as a basis for reminding.

Thus, the first stage of prompted recall may be the recognition of a recently experienced event (see Jenkins, Wald and Pittenger, in press). If this recognition proceeds on the basis of sufficient resemblance, not of identity, reminding itself can be considered a metaphoric process. The second stage of recall would be a process of regenerating the specific sentence constituents that originally led the comprehender to experience the event. The often regenerative nature of the second stage is evidenced by the kinds of paraphrases we cited above. This proposed model reverses the order of generation and recognition processes found in many two-stage models of recall (for example, Bahrick, 1970; Tulving and Thomson, 1973) and emphasizes the role of abstract relationships, rather than specific elements, as agents in the recognition phase. Considerable research is needed to determine the conditions under which recognition is likely to occur, and to differentiate between direct recognition of the earlier event (as in a déjà vu experience) and recognition mediated by some kind of search process. Subjects reported both types of recognition experience.

It is difficult to determine what kinds of representation, if any, to attribute to the comprehender of a metaphor. In these experiments, the grounds were formulated as verbal predicates. Since these were effective

133

prompts, it is tempting to assume that they prompted recall by accessing similar representations created during acquisition. This approach would accept the common assumption that sentence meaning is coded internally by means of a predicate or propositional notation system. An alternative possibility is that sentence comprehension is not representationally mediated, but is a vicarious engagement of the processes underlying perception and action (see Werner and Kaplan, 1963; Arnheim, 1969; Gibson, 1971; Verbrugge, 1977). Our characterization of domains in terms of structural and transformational invariants is consistent with this proposed alternative. If the role of a verbal prompt is to allow the listener to re-experience (recognize) a relation experienced at acquisition, prompts specifying that relation in any modality should be effective, that is, the relations may be abstract with respect to medium (verbal, optical, acoustic), as well as specific contents (tree trunks, straws, hoses, pipes). While propositional projections of abstract relations have considerable heuristic value for theoreticians, attributing these representations to the comprehender may preclude successful explanation of plasticity in word use and the imaginal processes that underly comprehension. Further study of the conditions for successful recall of metaphors may help direct the current controversy over "mental representation" (see Pylyshyn, 1973; Shepard, 1975; Kosslyn and Pomerantz, 1977).

The formal proposition has, for too long, been taken as the prototypical linguistic form. It has shaped the way we define the problems of expression, comprehension, and representation. For example, in many psycholinguistic tasks, subjects are asked to judge the validity of propositions about the outside world or about an artificial "experimental world." The subjects usually cooperate by implicitly adopting the experimenter's constraints: they respond realistically, conventionally, and normatively. Little attention is given to the possibility that the propositions rejected as "false" might be valid in appropriate metaphoric contexts. Many linguists and psychologists have adopted a similar implicit standard when developing theories for interpreting "deviant" expressions: they have attempted to normalize such expressions into standard axiomatic form, so that the canons of verification and inference will apply. While these exercises have some value for purposes of traditional linguistic description, they are of doubtful value as a basis for a theory of creativity in language use. The metaphoric "speech act" invites cognitive processes distinct from those engaged in accessing and verifying facts. Metaphor invites pretending, imagining, reasoning by analogy; in its more powerful forms, it requests a perception of resemblances by means of an unconventional reshaping of identities. The study of metaphoric competence in adults challenges us not to limit these processes to the nursery room and the therapist's couch, but to see them as crucial phenomena in the psychology of everyday life.

REFERENCES

Anderson, J. R. and G. H. Bower. (1973) Human Associative Memory. (New York: Wiley).

Anderson, R. C. and A. Ortony. (1975) On putting apples into bottles--A problem of polysemy. Cognitive Psychology 7, 167-180.

Arieti, S. (1974) Interpretation of Schizophrenia. (2nd ed.) (New York: Basic Books).

Arnheim, R. (1969) *Visual Thinking*. (Berkeley: University of California Press).

Asch, S. (1955) On the use of metaphor in the description of persons. In *On Expressive Language*, ed. by H. Werner. (Worcester: Clark University Press).

Bahrick, H. P. (1970) Two-phase model for prompted recall. *Psychological Review 77*, 215-222.

Barclay, J. R., J. D. Bransford, J. J. Franks, N. S. McCarrell and K. E. Nitsch. (1974) Comprehension of semantic flexibility. *Journal of Verbal Learning and Verbal Behavior 13*, 471-481.

Bickerton, D. (1969) Prolegomena to a linguistic theory of metaphor. *Foundations of Language 4*, 34-52.

Black, M. (1962) *Models and Metaphors*. (Ithaca: Cornell University Press).

Blumenthal, A. L. (1967) Prompted recall of sentences. *Journal of Verbal Learning and Verbal Behavior 6*, 203-206.

Blumenthal, A. L. and R. Boakes. (1967) Prompted recall of sentences. *Journal of Verbal Learning and Verbal Behavior 6*, 674-676.

Bolinger, D. (1965) *The atomization of meaning. Language 41*, 555-573.

Bransford, J. D. and J. J. Franks. (1973) The abstraction of linguistic ideas: A review. *Cognition: International Journal of Cognitive Psychology 2*, 211-249.

Bransford, J. D. and N. S. McCarrell. (1974) A sketch of a cognitive approach to comprehension. In *Cognition and the Symbolic Processes*, ed. by W. Weimer and D. Palermo. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Bransford, J. D., N. S. McCarrell and K. E. Nitsch. (1976) Contexte, comprehension et flexibilite semantique: Quelques implications theoriques et methodologiques. In *La memorie semantique*, ed. by S. Ehrlich and E. Tulving. (Paris: Bulletin de Psychologie).

Brown, R. W., R. A. Leiter and D. C. Hildum. (1957) Metaphors from music criticism. *Journal of Abnormal and Social Psychology 54*, 347-352.

Bühler, K. (1908) Tatsachen und Probleme zu einer Psychologie der Denkvorgänge. III: Über Gedankenerinnerungen. *Archiv für die gesamte Psychologie 12*, 24-92.

Cassirer, E. (1923) *Substance and Function*. (Chicago: Open Court).

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. (Cambridge: M.I.T. Press).

Cohen, L. J. and A. Margalit. (1972) The role of inductive reasoning in the interpretation of metaphor. In *Semantics of Natural Language*, ed. by D. Davidson and G. Harman. (Dordrecht: D. Reidel).

Collins, A. M. and M. R. Quillian. (1972) Experiments on semantic memory and language comprehension. In *Cognition in Learning and Memory*, ed. by L. W. Gregg. (New York: Wiley).

Craik, F. I. M. and E. Tulving. (1975) Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General 104*, 268-294.

Epstein, W. (1972) Long term memory for sentences, anomalous sequences, and random strings. *American Journal of Psychology 85*, 21-30.

Freud, S. (1950) *The Interpretation of Dreams*, trans. by A. A. Brill. [New York: Modern Library. (Originally published, 1900)].

Gibson, J. J. (1971) The information available in pictures. *Leonardo 4*, 27-

35.

Gombrich, E. H. (1968) Meditations on a hobby horse, or the roots of artistic form. In Aspects of Form, ed. by L. L. Whyte. (Bloomington: Indiana University Press).

Gordon, W. J. J. (1961) Synectics: The Development of Creative Capacity. (New York: Harper).

Hawkes, T. (1972) Metaphor. (London: Methuen).

Hesse, M. B. (1966) Models and Analogies in Science. (Notre Dame: University of Notre Dame Press).

Honeck, R. P., P. Reichmann and R. R. Hoffman. (1975) Semantic memory for metaphor: The conceptual base hypothesis. Memory & Cognition 3, 409-415.

Jenkins, J. J. (1966) Meaningfulness and concepts: concepts and meaningfulness. In Analyses of Concept Learning, ed. by H. J. Klausmeier and C. W. Harris. (New York: Academic Press).

Jenkins, J. J. (1974) Remember that old theory of memory? Well, forget it! American Psychologist 29, 785-795.

Jenkins, J. J., J. Wald and J. B. Pittenger. (in press) Apprehending pictorial events: An instance of psychological cohesion. In Minnesota Studies in the Philosophy of Science, vol. 9, ed. by C. W. Savage. (Minneapolis: University of Minnesota Press).

Katz, J. J. and J. A. Fodor. (1963) The structure of a semantic theory. Language 39, 170-210.

Keeler, C. E. (1961) Apples of Immortality from the Cuna Tree of Life. (New York: Exposition Press).

Kintsch, W. (1972) Notes on the structure of semantic memory. In Organization of Memory, ed. by E. Tulving and W. Donaldson. (New York: Academic Press).

Kintsch, W. (1974) The Representation of Meaning In Memory. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Koen, F. (1965) An intraverbal explication of the nature of metaphor. Journal of Verbal Learning and Verbal Behavior 4, 129-133.

Kosslyn, S. M. and J. R. Pomerantz. (1977) Imagery, propositions, and the form of internal representations. Cognitive Psychology 9, 52-76.

Lakoff, G. (1972) Hedges: A study in meaning criteria and the logic of fuzzy concepts. Papers from the Eighth Regional Meeting, Chicago Linguistics Society. (Chicago: Department of Linguistics, University of Chicago).

Langer, S. K. (1967) Mind: An Essay on Human Feeling, vol. 1. (Baltimore: Johns Hopkins Press).

Leech, G. N. (1969) A Linguistic Guide to English Poetry. (London: Longman).

Malgady, R. C. and M. G. Johnson. (1976) Modifiers in metaphors: Effects of constituent phrase similarity on the interpretation of figurative sentences. Journal of Psycholinguistic Research 5, 43-53.

Marks, L. E. and G. A. Miller. (1964) The role of semantic and syntactic constraints in the memorization of English sentences. Journal of Verbal Learning and Verbal Behavior 3, 1-5.

Matthews, R. J. (1971) Concerning a linguistic theory of metaphor. Foundations of Language 7, 413-425.

Mowrer, O. H. (1954) The psychologist looks at language. American

Psychologist *9*, 660-694.

Norman, D. A. and D. E. Rumelhart, eds. (1975) *Explorations in Cognition*. (San Francisco: Freeman).

Osgood, C. E. (1953) *Method and Theory in Experimental Psychology*. (New York: Oxford University Press).

Osgood, C. E., G. Suci and P. Tannenbaum. (1957) *The Measurement of Meaning*. (Urbana: University of Illinois Press).

Perfetti, C. A. and S. R. Goldman. (1974) Thematization and sentence retrieval. *Journal of Verbal Learning and Verbal Behavior* *13*, 70-79.

Piaget, J. (1950) *The Psychology of Intelligence*, trans. by M. Piercy and D. E. Berlyne. [London: Routledge and Kegan Paul. (Originally published, 1947)].

Piaget, J. (1962) *Play, Dreams, and Imitation in Childhood*, trans. by C. Gattegno and F. M. Hodgson. [New York: Norton. (Originally published, 1945)].

Pollio, H. R., J. M. Barlow, H. J. Fine and M. R. Pollio. (1977) *Psychology and the Poetics of Growth*. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Pylyshyn, Z. (1973) What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin* *80*, 1-24.

Richards, I. A. (1936) *The Philosophy of Rhetoric*. (Oxford: Oxford University Press).

Richards, I. A. (1960) *Coleridge on Imagination*. (Bloomington: Indiana University Press).

Rosch, E. (1973) On the internal structure of perceptual and semantic categories. In *Cognitive Development and the Acquisition of Language*, ed. by T. E. Moore. (New York: Academic Press).

Shaw, R. E., M. McIntyre and W. Mace. (1974) The role of symmetry in event perception. In *Perception: Essays in Honor of James J. Gibson*, ed. by R. B. MacLeod and H. L. Pick. (Ithaca: Cornell University Press).

Shepard, R. N. (1975) Form, formation, and transformation of internal representations. In *Information Processing and Cognition: The Loyola Symposium*, ed. by R. Solso. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Skinner, B. F. (1957) *Verbal Behavior*. (New York: Appleton-Century-Crofts).

Smith, E. E., E. J. Shoben and L. J. Rips. (1974) Structure and process in semantic memory: A feature comparison model for semantic decisions. *Psychological Review* *81*, 214-241.

Steinberg, D. (1970) Analyticity, amphigory, and the semantic interpretation of sentences. *Journal of Verbal Learning and Verbal Behavior* *9*, 37-51.

Sternberg, R. J. (1977) *Intelligence, Information Processing, and Analogical Reasoning: The Componential Analysis of Human Abilities*. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Thomas, O. (1969) *Metaphor and Related Subjects*. (New York: Random House).

Thomson, D. M. and E. Tulving. (1970) Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology* *86*, 255-262.

Tulving, E. (1972) Episodic and semantic memory. In *Organization of Memory*, ed. by E. Tulving and W. Donaldson. (New York: Academic Press).

Tulving, E. and D. M. Thomson. (1973) Encoding specificity and retrieval processes in episodic memory. *Psychological Review* *80*, 352-373.

Turvey, M. T., R. E. Shaw and W. Mace. (in press) Issues in the theory of action. In *Attention and Performance VII*, ed. by J. Requin. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Verbrugge, R. R. (1977) Resemblances in language and perception. In *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, ed. by R. E. Shaw and J. D. Bransford. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Weimer, W. B. (1973) Psycholinguistics and Plato's paradoxes of the *Meno*. *American Psychologist* *28*, 15-33.

Weinreich, U. (1966) Explorations in semantic theory. In *Current Trends in Linguistics*, vol. 3, ed. by T. Sebeok. (The Hague: Mouton).

Werner, H. and B. Kaplan. (1963) *Symbol Formation: An Organismic-Developmental Approach to Language and the Expression of Thought*. (New York: Wiley).

Skill Acquisition: An Event Approach with Special Reference to Searching for the Optimum of a Function of Several Variables*

Carol A. Fowler† and Michael T. Turvey††

ABSTRACT

Our paper divides into three parts. The first is a roughly hewn statement of the general orientation we wish to take toward the problem of skill acquisition. The second part develops a level of analysis that, in our view, is optimal for the examination of the problem; essentially, it is an ecological level of analysis that promotes the event rather than the performer as the minimal system that will permit an adequate explanation of the regulation and acquisition of skilled activity. The principal claims of the first two parts are highlighted in the third and final part through a detailed examination of a specific but prototypical coordination problem, namely, the problem of how one learns optimally to constrain an aggregate of relatively independent muscles so as to regulate a simple change in a single variable.

MOTOR TASKS, ACQUISITION PROCESSES AND ACTORS: A GENERAL ORIENTATION

It is prudent to preface a theoretical analysis of learning by some general comments on what the incipient theorist takes to be the nature of tasks that are learned, the nature of the processes that support the learning and the nature of the agent doing the learning. In the vocabulary of Shaw and McIntyre (1974), those three topics refer, respectively, to the three primary analytic concepts of psychology, namely, the what, how and who concepts. One can argue that this set of analytic concepts is closed, that is, that the concepts are logically co-implicative (Shaw and McIntyre, 1974; Turvey and Prindle, 1978). The closure of the set is illustrated by the following example (Shaw and McIntyre, 1974):

---

*To appear in Information Processing in Motor Control and Learning, ed. by G. Stelmach. (New York: Academic Press).

†Also, Dartmouth College, Hanover, New Hampshire.

††Also, University of Connecticut, Storrs.

Acknowledgment: This work was supported by NIH Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 1 (1978)]

> The degree of hardness of a sheet of metal tells us something about the nature of the saw we must use to cut it (i.e., something about what is to be done); a blueprint or pattern must be selected in the light of what can be cut from the materials with a given degree of tolerance (i.e., how it is to be done); while both of these factors must enter into our equations to determine the amount of work that must be done to complete the job within a reasonable amount of time. This latter information provides a job description that hopefully gets an equivalence class of existing machines rather than a class that might accomplish the feat in principle but not in practice (i.e., implies the nature of the who or what required to do the task). (p. 311)

A Parallel Between Evolution and Learning

In search of a general orientation to the nature of tasks, processes and agents as they bear on the issue of skill acquisition, we are drawn to the parallel between a species participating in the slow process of evolution and an individual animal participating in the comparatively rapid process of learning.

From a perspective that encompasses the whole evolving world of living systems, any given species appears to be a "special purpose device" whose salient properties are those that distinguish the given species from other species. These salient properties, synchronically described, mark the state of adaptation of the species to the special and relatively invariant properties of its environment. In the course of time, the species maintains its special attunement by coupling its evolution to that of its changing environment.

If the perspective is considerably narrower, encompassing only the lifetime and habitat of an individual animal, then the system being observed appears to be a "general purpose device" to the extent that the individual animal can enter into various temporary relationships with its environment. In the course of ontogeny, the individual animal adds to its repertoire of skilled acts.

It is roughly apparent that the "evolution" in ontogeny of a skilled act parallels the evolution of a species. Adaptation to an environment is synonymous with the evolution of special biological and behavioral features that are compatible (symmetrical) with special features of the environment. Similarly, we may claim that facility with a skill is synonymous with the ontogeny of special coordinative features that are compatible with the special features of the skill. Insofar as an environment has structure that provides the criteria for adaptation, so we may expect, not surprisingly, a task to have structure that provides the source of constraint on skilled solutions. Insofar as a species is said to be a particular biological attunement to a particular niche, we may wish to say, perhaps curiously, that the individual animal, as skilled performer, is a particular attunement to the particular task that it performs skillfully. This last and cryptic parallel must be commented on further, for aside from requiring clarification, it contains within it a potentially useful metaphor for the understanding of coordinated activity.

Consider the proposition that an animal and its environment are not logically separable, that one always implies the other. An animal's environment should not be construed in terms of the variables of physics as we commonly understand them; a considerably more useful conception is in terms of affordances (Gibson, 1977). An affordance is not easily defined, but the following may be taken as a working approximation: "The affordance of anything is a specific combination of the properties of its substance and its surfaces taken with reference to an animal" (Gibson, 1977, p. 67). Thus, for example, the combination of the surface and substance properties of rigidity, levelness, flatness and extendedness identifies a surface of support for the upright posture and locomotory activity of humans. Put another way, an object or situation, as an invariant combination of surface variables, affords a certain activity for a given animal if, and only if, there is a mutual compatability between the animal, on the one hand, and the object or situation on the other.

Affordances are the aspects of the world to which adaptations occur. Consequently, we can now identify the special features of the environment referred to above as a set of affordances, equate a "set of affordances" with a "niche" (Gibson, 1977) and recognize that a set of affordances is perceptually and behaviorally occupied by an animal. It is in this sense that an animal and an environment are not logically separable; for a niche implies a particular kind of animal and a species implies a particular kind of niche (Gibson, 1977).

A crude but useful metaphor is that the fit between an animal and its niche is like the fit between the pieces of a jigsaw puzzle. Figure 1 depicts the fit for a minimally complex puzzle. Following the jigsaw puzzle metaphor, adaptation and attunement are synonyms for the fit of a species to a niche. It is in this same metaphorical sense that skill acquisition can be understood as attunement: in terms of a two-piece jigsaw puzzle, one piece is an appropriate dynamical description of the skill and the other piece is an appropriate and complementary dynamical description of the animal.

The Actor as a Mimicking Automaton

To pursue further the idea of skill acquisition as attunement, let us return to the notion of the individual animal as a general purpose device. The animal of interest to us is, of course, human. In deliberations on perception, the human is often referred to as the perceiver; in deliberations on action, therefore, it seems appropriate to refer to the human as the actor.

We wish to claim that the individual actor is a general purpose device, not because he or she has the capacity to apply a single, general purpose action strategy to the skill problems encountered, but because he or she has the capacity to become a variety of special purpose devices, that is, a variety of specific automata.[1] The distinction between these two kinds of general purpose devices is depicted crudely in Figure 2. One device can accept only one program and generalizes that program across a variety of tasks. The other device can accept a variety of programs, one program for each of a variety of tasks. The familiar paradigm for learning theory, associationism, identifies the actor as a general purpose device of the first kind. It can be shown that a formal statement of associationism, the Terminal Meta-Postulate (Bever, Fodor and Garrett, 1968), is formally equivalent to a strictly finite state automaton that accepts only one-sided (right or left) linear grammars (Suppes, 1969). Such an automaton is formally incapable of natural language and complex coordinated movements, to name but a few limitations. A person, on the other hand, is obviously capable of such things and more besides. Nevertheless, it is reasonably fair to claim that, on the grounds of mortality and finite computing capacity, our actor, a person, is a machine with finite states. How then does he behave as if he were a machine of a more powerful kind, such as a linear-bounded automaton that accepts context-sensitive grammars? One hypothesis (Shaw, Halwes and Jenkins, 1966) is that the class of finite state machines that best characterizes the individual person is that of finite state transducers. These machines transduce the behavior of more powerful machines into equivalent finite state behaviors; they are capable of processing the same inputs as more powerful machines, but only up to some finite limit. In short, the individual actor as a finite state transducer can "mimic" the competency of more powerful automata, that is to say, he or she can become, within limits, any one of a variety of special purpose devices whose complexity is compatible with the complexity of the task it must perform.

---

[1] Turvey, Shaw and Mace (in press) have introduced a similar distinction between "hierarchies" and "coalitions." In the context of the present discussion, a hierarchy is a general-purpose device of the first type and a coalition a general-purpose device of the second type.

We do not wish to push the interpretation of the actor as finite state transducer too far. We wish to view it more as an analogy, for there are reasons to believe that the general machine conception, of which finite state transducers and the like are examples, may well be inappropriate for biology.[2] Nevertheless, the preceding is sufficiently instructive for our current purposes: it identifies our general orientation to the agent--that is, the actor--as a mimicking automaton. We can now make a further comment on the idea of skill acquisition as attunement: it is, in large part, the idea that an actor becomes that particular kind of machine that is consonant with the essential feature of the particular skill that the actor is performing.

Summary

We summarize these prefatory remarks with a tentative answer to the question: What is it about an actor and about the skills that he seeks to perform that he can (learn to) make of himself a variety of special purpose devices? First, in reference to the nature of the actor: the relationships among muscles are sufficiently plastic so that within limits, actors are able to constrain or organize their musculature into different systems. From this perspective, learning a skill involves discovering an optimal self-organization. Second, in reference to the nature of skills: skills have structure, and discovering an optimal self-organization is in reference to those variables of stimulation corresponding to environmental and biokinematic relations that specify the essential features of the skill the actor is to perform. This raises the important question of what are the useful skill-specific variables of stimulation that, in the course of acquiring a skill, guide and regulate the current approximation and prescribe the next approximation to the desired performance (attunement). Third, in reference to the nature of the processes supporting learning: insofar as the useful skill-related information must be discovered, the actor must engage certain "search methods" that reveal that useful information to him. These search methods must be compatible with the actor, that is to say, they must be compatible with, for example, real-world mechanical and temporal constraints that natural (as opposed to abstract) actors must obey.

---

[2] Shaw, R., T. Halwes and J. Jenkins. (1966) The organism as a mimicking automaton. (Unpublished manuscript, Center for Research in Human Learning, University of Minnesota).

DEFINING THE DOMAIN OF SKILL ACQUISITION FOR A THEORIST

In seeking an explanation of anything, it is important that the forms of theoretical and investigatory attention be a domain of entities and functions that is optimal to the particular problem under investigation. "Optimal domain" means two things. First, any decision to investigate a problem involves selecting some system (some collective of entities and functions) as the minimal one that is relevant to the problem's explanation. If the selected system excludes some entities and functions that are, in fact, crucial to the explanation, they exert an influence on the selected system that, from the observer's perspective, is random (see Bohm, 1957). In consequence, the system's behavior to those perturbations may be inexplicable.

Equally important is the second sense of "optimal domain." Any given system may be described at several different levels where each level is distinguished by the entities and functions to which its vocabulary refers. Importantly, different levels of description of a system make available to the theorist different concepts that he can invoke in his explanation (Medawar, 1973; Putnam, 1973). Which concepts are more useful to the theorist depend on what problem he has elected to explain.

What should be the minimal system for a theory of the acquisition and performance of skilled activity? At first blush, the actor looks to be the appropriate unit deserving observation and systematic measurement. With the actor as the minimal system, the concept of coordination can be judiciously defined in terms of relationships defined over the muscles and joints of the body. The locus of movement control can be given relatively precise coordinates, namely, the nervous system of the actor. However, in taking the actor as the minimal system, we adopt a myopic view of the contribution of the environment to coordinated activity. This is not to say that an actor-oriented approach to the theory rejects the environment's contribution, but rather that it detracts from a serious analysis of the environment as the necessary support for coordinated, skilled movements. An actor-oriented perspective on skill, with its pinpointing of the actor as the source of control, encourages the impoverished description of information about the environment as sensory signals whose meaning is contributed wholly by the actor (see Schmidt, 1975).

The claim we wish to make is that a superordinate system, one that encompasses the actor, his actions and the environmental support for his actions, is the minimal system whose observation will permit an adequate explanation of the regulation and acquisition of skilled performance. To anticipate, this minimal system will be referred to as an event. From the perspective of this system, coordination is a relation defined over the actor and the environment, and control is the exclusive prerogative of neither.

What should be the level of description for this minimal system? Putatively, the theorist who aims to explain the acquisition and performance of skilled activities should select a level of description that is compatible with an actor's own self-description and with the actor's descriptions of the environment. The theorist should select a grain-size vocabulary that, in reference to skilled activity, includes those entities and functions that are regulated by actors and those entities and functions that are regulative of actors.

Our previous discussions of coordinated movement (Fowler, 1977; Turvey, 1977b; Turvey, Shaw and Mace, in press) may be characterized as attempts to select and define an appropriate level of description of acting animals and of the environments in which they act. We will summarize and elaborate on those attempts in the remarks that follow.

Events as Significant Units of Observation in a Theory of Skilled Action

An act performed in a natural context has two sources of control: one is the actor himself, and the other is the environment in which the act occurs.

To achieve some aim, whatever it may be, an actor engages in a systemic relationship with the environment. That is, he regulates his body in relation to environmental sources of control such as gravitational and frictional forces. His task, then, is quite different from one of producing an act *in vacuo*; it is to generate a set of forces that, *together with* the environmental forces impinging on him, are sufficient to achieve his aim. In the sense of the jigsaw puzzle metaphor, the forces supplied by the actor complement those supplied by the environment. Furthermore, the actor's aim itself is not entirely a product of his own will. Rather, it must be some selection on his part among the limited possibilities afforded by the environment.

In short, we can say that actors and their environments participate in a larger system that we will call an "event," following the usage of Shaw, McIntyre and Mace (1974). Structurally described, an event includes the actor and the environmental support for his actions. "Environmental support" includes the surfaces, objects and living systems in relation to which the actor governs his behavior and, in addition, the structured media (such as the ambient light and air) that provide the actor with an event's functional description--that is, with a specification of what is happening in the course of an act.

Two principles derive from the foregoing discussion. First, an actor controls the functional description of an event rather than the functional description of his own body; and second, an appropriate observational perspective of a theorist of skilled action is a perspective that encompasses events rather than actors only. The two principles are illustrated in the following example.

Consider a person changing a flat tire on his car. The tire-changing event includes the actor's removing the spare tire and jack from the trunk of his car, jacking up the car and replacing the flat tire with the spare. The actor's movements in the course of the tire-changing event and his (inferred) self-commands to movement have no apparent rationale if they are observed in isolation. For instance, the rhythmic up and down gestures of the actor's arms during one phase of the event may be rationalized by an observer only if he recognizes that the arms are operating the handle of the jack and that the flat tire is being raised off the ground.

More than simply controlling his own movements, an actor controls the character of the event in which one of the participants is himself and the other is the environment. He deems his performance successful if he imposes his intentions on the character of the event. Put another way, an actor has achieved his aim if an observer's description of the event in which the actor participates is synonymous with the actor's description of his intentions.

In sum, an appropriate observational perspective for a theorist includes both the actor and the environment in which he acts. A more limited perspective that excludes or minimizes the environment is likely to remove the means by which an observer can either detect the actor's intent or rationalize aspects of his performance.

An Appropriate Level of Description of Events, Actors and Environments

Events have been promoted as the minimal systems to be observed for the development of an adequate theory of skilled action. Primarily, the grounds for this selection are that no systems smaller than events encompass those entities and functions over which actors exert their control. The same kind of selection criterion may be invoked in a choice of "level of description." Having selected an observational unit, it is necessary to choose a descriptive vocabulary for it. Again, it seems most appropriate to select a grain-size of vocabulary such that its referent entities and functions are those that populate the actor's habitat from his observational perspective, because those are the things with which he deals in the course of his actions.

In the next sections we will select a level of description of an actor and of his habitat. In the case of an actor, our aim is to select a vocabulary that mimics the effective *self*-descriptions putatively invoked by actors as a means of controlling their actions. Similarly, our aim is to select a level of description of the environmental media that is isomorphic with the grain-size of the information detected by actors. Hypothetically, a description of the structured media that captures the significant information for actors is concomitantly a description of the environmental entities and functions that, from the actor's perspective, constitute his habitat (see Shaw et al., 1974; Gibson, 1977).

The Actor. An actor can be described exhaustively in several ways where each "way" is defined by the primitive entities to which its vocabulary refers. These ways are significantly restricted if we assume that the aim of a theory of coordinated activity is to specify what an actor controls when he performs an act. In this respect, it is not suprising that no one has ever devised a theory of coordinated activity in which the primitive units of vocabulary are the individual cells or molecules of the actor's body.

Presumably, two reasons why neither cells nor molecules have been proposed as the primitive entities of a theory of action are, on the one hand, that an actor could not possibly control those microscopic entities and, on the other hand, that even if he could, he would not choose to do so. For each cell whose trajectory he wished to control, an actor would have to provide values for as many as six degrees of freedom.[3] It is inconceivable that he could continuously set and reset the values of the six degrees of freedom of the millions of cells whose state trajectories are regulated in the course of an act.

Even if he could control that many degrees of freedom, to do so would constitute a gross violation of a principle of least effort. The cells in the actor's body are constrained to act as systems of cells. The degrees of freedom of these collectives are orders of magnitude fewer than the summed degrees of freedom of the individual cells in the collectives. A more abstract level of description of an actor than one whose primitive entities are cells, captures these constraints on classes of cells by treating each class or collective as an irreducible unit. Thus "deltoid muscle" refers to a collective of cells that are constrained to act as a unit.

If an actor exploits an abstact level of self-description on which muscles are irreducible units, he indirectly takes care of the vast multitudes of degrees of freedom of his individual cells by directly controlling the many fewer degrees of freedom of collectives of cells.

What is more, the "muscular" level of description is less powerful, but in a useful way, than a microscopic level. If an actor were to control his individual cells directly, he would specify values for their trajectories that he could never achieve because they violate the constraints on collectives of cells (for example, the combined trajectories might entail the disintegration of a muscle). In order to preclude such violations, the actor would have to know a set of rules for combining cell trajectories. However, he can avoid knowing anything about these rules if he selects a more abstract way of describing himself.

We have belabored the obvious point that actors control larger entities than cells and molecules in order to bring out some reasons why one level of description of an actor may be more useful to a theorist than another. Let us summarize these arguments before suggesting a less obvious point--that a level of description on which muscles are the irreducible units may not be sufficiently coarse-grained to be useful either to an actor or to a theorist.

---

[3] The six degrees of freedom are the values of the instantaneous positions and velocities of a cell on each of the three spatial coordinate axes.

Some levels of self-description are impossible for an actor to use because they demand that he provide values for vast numbers of degrees of freedom. Relatively macroscopic or abstract levels of self-description help to solve the "degree of freedom problem" (see Turvey et al., in press) by classifying the entities of the microscopic level and hence their degrees of freedom. The abstract levels provide one label for large numbers of elementary units that are constrained to act as a collective. By controlling the few degrees of freedom of the collective, the actor thereby regulates the many degrees of freedom of the components. The more abstract description is the less powerful one, but it is less powerful in a useful way. It allows the actor to know less of the details of the system that he controls, but to regulate it more easily and effectively (see Greene, 1969, 1972). Finally, concepts emerge (for example, "muscles") at a macroscopic level of description that do not exist on microscopic levels because the concepts refer to constraints on, or patternings of, entities that are treated as individuals on a microscopic level (see Medawar, 1973; Putnam, 1973).

Several theorists and investigators have proposed that an actor controls *groups* of muscles rather than individual muscles (for example, Weiss, 1941; Easton, 1972; Turvey, 1977b). Their reasons for preferring the more abstract description of an actor are those given above. An actor *cannot* govern his muscles individually because to specify values for their total number of degrees of freedom would be impractical if relevant cost variables are considered (Shaw and McIntyre, 1974; Turvey et al., in press). Greene (1969) estimates that there are over forty degrees of freedom in the hand, arm and shoulder alone, and dozens more in the trunk, shoulders and neck. Furthermore, the relationships between a central command to a muscle, the muscle's behavior and the movements of a limb are indeterminate both physiologically and mechanically (see Hubbard, 1960; Bernstein, 1967; Grillner, 1975; Turvey, 1977b). Commands to individual muscles would appear to constitute an inappropriate vocabulary of control for an actor.

Yet, even if an actor could control his individual muscles, there are reasons for believing that he would not choose to do so. First, the actor's muscles are organized into functional collectives. Some collectives, the reflexes, appear to be "prefabricated" (Easton, 1972). However, many--those involved in locomotion for instance (for example, Grillner, 1975; Shik and Orlovskii, 1976)--are marshalled temporarily and expressly for the purpose of performing a particular act. There is ample evidence that these systems of muscles that we have called "coordinative structures" (Fowler, 1977; Turvey, 1977b; Turvey et al., in press) after Easton (1972), are invoked by actors in the performance of large varieties of acts (for example, speech: see Fowler, 1977, for a review; locomotion: see Grillner, 1975, for a review; swallowing, chewing: Doty, 1968; Sessle and Hannam, 1975). The actor's organization of his musculature into coordinative structures that are especially appropriate to the performance of a *limited* class of acts is what we mean when we describe an organism as a general-purpose device by virtue of its capacity to become a variety of special purpose devices.

The constraints on groups of muscles that organize them into collectives are different in kind from those on some groups of cells, for instance those that constitute a bone and perhaps those that constitute a muscle. The label "bone" refers to a group of cells constrained to adopt a particular macroscopic form. It seems clear in this case that the constraints have exhausted the *configurational* degrees of freedom of those cells. The result is a rigid body. In contrast, the constraints that yield a coordinative structure appear to be a kind that Pattee (1973) calls control constraints. Control constraints, like structural constraints, are classifications of the degree of freedom of elementary components of a system, but they regulate the *trajectories* of a system rather than its configuration. Hence, a coordinative structure is a four-dimensional system that may be identified by what it *does*.

If the actor's vocabulary of self-description or self-control refers to coordinative structures rather than muscles or, equivalently, if it refers to the control constraints on this musculature, then apparently his descriptions are functional in nature.

A level of self-description in which the coordinative structure constitutes the elemental unit of vocabulary is less powerful than one in which muscles are described but, again, the loss of power is beneficial to the actor. If muscles are the primitive units of description for the actor, then he can prescribe combinations of muscle contractions that never occur because they violate the constraints on groups of muscles. In the terms of Weiss (1941), the too-microscopic level of description cannot explain why actors limit themselves to coordinated movements and avoid "unorganized convulsions." The macroscopic level allows an actor to exploit the constraints on groups of muscles that putatively limit him to performing coordinated movements.

Finally, on the coarse-grained level of description, concepts or properties emerge (for example, in coordinative structures) that do not exist on the more detailed levels of description. These concepts or properties derive from the constraints on the individual elements of those detailed levels. For instance, the coordinative structures are nested. This property is well-documented again for the relatively simple act of locomotion (for example, Easton, 1972; Grillner, 1975). Each coordinative structure governs an activity. A nested set of coordinative structures may govern a long sequence of movements with little detailed executive control being required of the actor. In fact, the sequence of autonomously generated movements may be indefinitely long as in walking or chewing or breathing, if the "repertoire" of the nested coordinative structures regenerates itself cyclically (see Fowler, 1977).

Since many of the coordinative structures are not "prefabricated," the problem for an actor is to marshall those groups of muscles that will accomplish his purposes. The view of an actor provided by a coarse-grained description of him suggests the forming of relevant coordinative structures as a primary problem of skill acquisition.

The Environment in Relation to an Actor

Environmental Affordances. A component of an environment populates an actor's world only if the actor can engage in some relationship with it that has significance for him. More simply, the meaning of the component for an actor is captured by specifying the set of events in which the actor and component may participate (see Sperry, 1952; Shaw et al., 1974; Gibson, 1977). These potential relationships between actors and environment-components are what we called earlier the "affordances" of the components for the actor.

We can provide a different perspective on the concept of "affordance" by reexamining the nature of an event. The character of an event, in particular its functional description, is determined by the totality of forces exerted by and on the various event-participants. Among the forces that shape the character of an event are gravitational forces, which are extrinsic to the actor, and frictional and contact forces, which are generated by the actor's encounter with the environment. In addition to these, are the forces that enable an actor more directly to regulate the character of an event. They are the forces generated by the actor's own muscular activity.

Clearly, actors cannot achieve an aim to perform an act by generating all of the forces necessary to get the job done. Rather, they must contribute to the totality of extant forces just those muscular forces that will *bend* the character of an event in the desired direction.

By hypothesis, the affordances of an environment for an actor, as given in the structured environmental media, are the sets of forces (of adaptive significance to him) that the actor can generate in collaboration with the extant forces, and the relation to the environment. The totality of forces that the actor selects from among the potential ones defines his intent. For a skilled actor, the intent becomes, through his muscular efforts, the functional description of the event.

The Structured Media. The structured media, that is, the ambient light and air, etc., apprise actors of the properties of an event; they are said to contain *information about* events in the sense of *specificity to* events.

The media are components of an environment that, relative to other components, are compliant. Thus, for example, when light contacts some surface, the light but not the surface is significantly altered. In particular, the amounts of light reflected from a surface in a given direction and the wavelengths of the light are specific to various properties of the surfaces; the slant of the surface relative to the source of radiant light, its composition and so on. Hence the light, on contact with the surface, is *constrained* (or is patterned) in its subsequent behavior by the properties of the surface. Furthermore, the patterning of the rays of light is *specific* to the source of its patterning. Therefore, the structure in the light is isomorphic, though abstractly so, with the properties of the structure's source. Just as an environment is constituted of nestings of entities and functions, a medium contains structure of various grain-sizes. However, the structure of interest to an actor and to a theorist is only that which is specific to, or isomorphic with, the properties of the *event* in which the actor is participating. The environmental entities and functions that are specified *to an actor* by the structure of a medium are just those whose properties are of adaptive significance to him.

We believe that this is a crucial observation. The light to an eye is amenable, as is the actor himself, to various levels of description (see Mace, 1977). Typically, as Gibson has noted (for example, 1961), theorists take as their unit of description the individual ray of light that has only the properties of wavelength and intensity. The individual rays are meaningless to an actor; pursued through his nervous system, they excite receptors on the retina and are transformed into still-meaningless "raw" sensory signals (for example, Schmidt, 1975). They are supposed to acquire significance only as the actor learns to assign meaning to them via the efforts of his community of coactors who provide him with "knowledge of results."

This view is fostered by a too microscopic level of description of the light and of its neural consequences. In particular, it is too fine-grained to represent what in the light is genuinely informative and significant to an actor, just as the levels of description of an actor in which cells or muscles are the descriptive units are too fine-grained to capture the properties of the muscle systems that actors exploit. That level of description of the light that considers only two variables (intensity and wavelength) fails to capture any of the constraints on the paths, spectral compositions and intensities of bundles of light rays that are specific to (and hence that specify to a perceiver) the environmental sources of the constraints. In contrast, if the sensitivity of perceptual systems is not to the microscopic properties of a structured medium, but rather to the constraints or to the structure itself--that is, to a macroscopic level of description of the medium--then actors need not learn to manufacture a significance for stimulation. The meaning or significance is the set of properties in the environment that structured the light and therefore, that are specified by it with reference to an actor.

Other investigators have cataloged some of the information in the structured light available to an actor (for example, Gibson, 1958, 1961, 1966, 1968; Lee, 1974, 1976; Turvey, 1975, 1977a, 1977b). Here we provide only a brief description, but one that is sufficient for our later consideration of the role of higher-order variables of stimulation in the control and acquisition of skilled acts.

The patterning of the ambient light to an eye provides an actor with information about: (1) the layout of environmental surfaces and objects, (2) what is happening in the course of an event, (3) what is about to happen and when it will occur, and (4) the possibilities for control by the actor over what happens. We will consider each in turn.

Information About Layout Provided at a Stationary Point of Observation. The optic array is the set of light rays that reflect off of environmental surfaces and converge at all possible points of observation in the environment (Gibson, 1961). The portions of the array that converge at a single point of observation may be described as a nested set of "visual solid angles"[4]. A visual solid angle is a closed sector of the array with its apex at the point of observation. It is set off from its neighboring angles by differences from them in the intensity and spectral composition of its component rays by light. Each visual solid angle corresponds to a component of the environment where a component may differ from its neighbors in shape, slant relative to the source of illumination, distance from the observer, and properties of its material composition that determine its spectral and nonspectral reflectance.

---

[4] Gibson, J. J. (1972) On the concept of the "Visual Solid Angle" in an optic array and its history. (Unpublished manuscript, Cornell University).

Some properties of the environmental correlates of a visual solid angle are specified by the angle's cross-sectional shape, its intensity, and its spectral composition. The borders of an angle typically correspond to the edges of an object in the environment.

Visual solid angles are nested because environmental surfaces and objects are textured. That is, the structure of an environmental surface or object is specified by a corresponding patterning of visual solid angles in the optic array.

More information about structure, as well as information about change, is given in a transforming, rather than a static, optic array.

The Structural and Functional Descriptions of Events Given by a Transforming Optic Array. According to Pittenger and Shaw (1975), two kinds of information exhaust the information-types provided by the structured media of an event. A _structural invariant_ is information about shape or, more accurately, about persistent identity that is preserved across (physical) transformation. A _transformational invariant_ is information about physical change that is preserved across the different structures that may support the change. (See also Turvey, 1977a). These two kinds of information provide an actor with an event's structural and functional descriptions.

As an actor moves through an environment, he continually changes his observational perspective of it. If (solely for convenience) we describe this continuous change of perspective as a succession of discrete changes, we may say that the moving observer successively intercepts new observation points as he moves. The optic array at each of these fictitiously abstracted observation points constitutes information about layout of the sort described in the preceding section. The information at one observation point may or may not be sufficient to specify unambiguously to an observer the layout of environmental surfaces and other components relative to him. However, there is only one environmental layout that is consistently possible _across_ a set of connected observation points (Gibson, 1966). More accurately, the layout of environmental surfaces that is given in a transforming optic array is just that one layout whose persistent identity is specified throughout the transformation.

A global transformation of the optic array is effected when an actor changes his perspective on the environment. What is invariant (or what has persistent identity) across perspectives is the environmental layout. What changes with the observation point is information about the actor's perspective on the environment. That is, a global transformation of the optical structure is effected by the actor's movements and continually provides information on his relationship to the components of the environment. In short, global transformations of the optic array are specific to an observer and to his path through the environment (Lishman and Lee, 1973; Lee and Aaronson, 1974; Lee, 1976; Warren, 1976).

Now consider object motion from a stationary perspective. As an object in the environment changes its location relative to a stationary point of observation, its corresponding visual solid angle in the optic array undergoes transformation. The nature of the changing relationship between observer and observed is specified, in part, by the nature of the angle's transformation (that is, by the symmetrical or asymmetrical magnification or minification of the angle's cross-sectional area). More than this, it is also specified by the angle's progressive occlusion and disocclusion of those components of the optical structure that correspond to foreground and background components of the environment (Gibson, 1968).

For example, as an object approaches an observer head on, the cross-sectional area of the corresponding visual solid angle at the place of observation expands symmetrically. The bottom or leading edge of the angle progressively occludes foreground optical texture, while the top, or trailing edge, disoccludes the optical texture corresponding to the object's background. The lateral edges effect a shearing of optical texture.

Both kinds of transformation (that is, symmetrical magnification of a visual solid angle; occlusion, disocclusion and shearing of optical texture) specify motion in a restricted part of the environment and, in the absence of additional information that the actor is pulling the object towards him, specify motion due to forces extrinsic to the actor.

The Specification of Future Events. If an actor approaches a barrier or other object head on, the visual solid angle corresponding to it undergoes symmetrical magnification. Its rate of magnification specifies the actor's rate of approach. The fact that the magnification is symmetrical indicates to an appropriately attuned actor that he will collide with the barrier if the current inertial conditions continue. (A nonsymmetrical expansion indicates, depending on the degree of asymmetry, that the actor will bypass the barrier or that he will collide with it to the left or right of its center.) More than the fact of imminent collision, Schiff (1965) and Lee (1974, 1976) show that the time-to-collision is also specified to an observer by the transforming optical structure.

Thus, the macroscopic patterning of the transforming optic array provides the actor with information about what is currently happening and with information about what will happen if the current conditions persist (see Lee, 1976).

The Affordance Structure of Events. Of major important to an actor attempting to impose his intentions on the character of an event is information that prescribes to him the directions in which his contributions of muscular force can alter the current inertial conditions. To take a simple example: when we say that a surface affords locomotion for an actor, we mean, in part, that the ambient light (or some other structured medium) specifies to the actor the nature of the reactive forces (the frictional and contact forces) that the surface will supply, given his attempts to walk on it.

Information about the rigidity of a surface and about its slant and composition is concomitantly information about the surface's potential to participate in an event that includes the actor's walking on it.

This information is only information about walk-on-ability in relation to additional information about the actor's somatotype, however. That is, the affordances of a surface (or object) are the events in which the surface and the actor may participate, and they are contingent on the properties of the surface considered not absolutely, but relative, to properties of an actor. Hence, to detect the affordances of an environment-component, the actor has to detect _body-scaled_ information--that is, information about the component's properties relative to his own.

Lee's (1974) analysis of the optical information available to a locomoting observer indicates that information about the position coordinates of objects in the environment and information about the actor's rate and acceleration of movement are provided in units of the observer's own height. Is it possible that information about the actor's general build and perhaps, therefore, about his potential to contribute to the forces governing an event is provided in global transformations of the optic array? When he is walking, there are global transformations due to his sinusoidally shifting center of gravity. The extent of shift in the left-right and up-down directions as well as in the direction of walking may correlate with an actor's size and weight.

These shifts in the center of gravity effect rhythmic changes in the horizontal and vertical distance of the actor's head from components of the ground plane. Hence, the actor effects a transformation of optical structure that is specific to his rhythmically changing perspective on the environment. If the transformation in turn is specific to the actor's somatotype, it also provides information about his potential to contribute muscular force to an event.

Concluding Remarks: Increasing Controllable Degrees of Freedom so as to Secure Certain Reactive Forces

We began by selecting an observational domain for a theory of skilled action that we labeled an "event." We considered events to be the minimal observational domains that include, on the one hand, all of the entities and functions over which actors exert their control and, on the other hand, the entities and functions that are regulative of actors. Following that, we selected compatible descriptive vocabularies for the different components of an event. Our selections are more coarse-grained than the vocabularies typically adopted by theorists of skilled action. However, we defended them on the grounds that it is precisely the _patternings over_ microscopic entities and functions that are signified to actors and not the microscopic components themselves.

Our method of selecting the descriptive vocabularies was one that fractionated the event into its components. We will conclude this section of the paper by reconstructing the event concept and by describing one way in which it enriches a developing theory of skilled action and skill acquisition.

One orientation to coordinated activity, as cited above, is that acts are produced through the fitting together of autonomous subsystems (coordinative structures), each of which "solves" a limited aspect of the action problem. In this orientation, the actor's plan, that is, his abstract self-description, is regarded as the specification of that which remains when the contribution of the autonomous subsystems is subtracted out. The action plan supplies the coordination that is not supplied by the coordinative structures.

Precisely what is it that coordinative structures supply? One answer might be that they autonomously supply certain relations among various parts of the body. The difficulty with this answer is that, left unqualified, it steers dangerously close to an "Air Theory" formulation (see Gibson, 1950) of coordinated activity in which the actor, for all intents and purposes, is construed as suspended in a vacuum oblivious to external environmental forces. An "Air Theory" formulation speaks more to the mining of coordinated activity than to coordinated activity itself, for coordinated activity requires environmental support for its proper functioning.

Necessarily, an event perspective expresses the contribution of the environment to coordination. Coordination in the event perspective is defined not in terms of biokinematic relationships (that would be so if the actor were taken as the unit of analysis), but in terms of relationships among forces, those forces supplied muscularly by the actor and those supplied reactively and otherwise by the environment. The surfaces of support, the participating structures (such as other actors, striking implements, etc.), the biokinematic links and gravity, provide the actor with a large potential of reactive forces. This emphasis on what the environment provides characterizes the event perspective as a "Ground Theory" formulation of coordinated activity: an activity cannot logically be separated from its environmental support.

Consider environmental surfaces. These afford reactive forces that are opposite and approximately equal (although not always equal; it depends on the composition of the surface) to the forces generated by muscle activity. Thus in walking, the actor secures by his muscular efforts reactive forces that propel the body forward at one moment and restrain the forward motion of the trunk at the next. In leaping a high barrier, the actor applies his muscular forces in such a fashion as to secure reactive forces that are more nearly vertical than horizontal.

Of course, when the actor is not in contact with a supporting surface but is moving in the air, then the equal and opposite reaction to a motion of parts of the body occurs within the body itself. Swiftly moving the arm at shoulder level from a sideward to a forward position will rotate the body about its longitudinal axis in the direction of the moving arm. This aside bears significantly on the contrast between the actor/air theory formulation and the event/ground theory formulation in that the same movement performed when the body is in the air and when it is in contact with a rigid surface secures very different reactive forces with very different coordinative consequences.

Consider biokinematic chains. These obey the principles of kinematic chains in general; for example, a controlled movement of one link of the chain

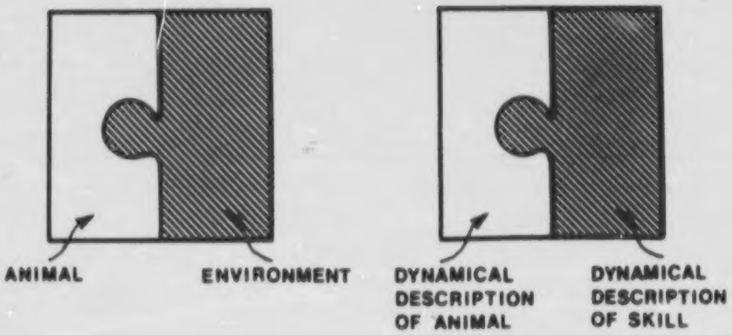

ANIMAL
ENVIRONMENT
DYNAMICAL DESCRIPTION OF ANIMAL
DYNAMICAL DESCRIPTION OF SKILL

Figure 1: The jigsaw puzzle metaphor.

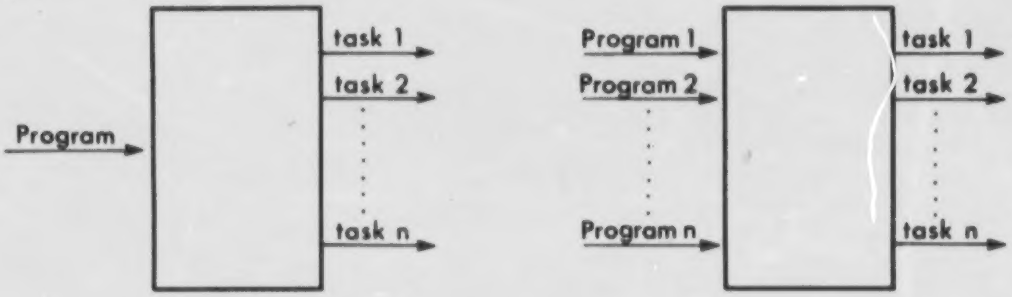

Program
task 1
task 2
task n
Program 1
Program 2
Program n
task 1
task 2
task n

Figure 2: Two kinds of general purpose devices. The one on the left accepts only one program and generalizes that program across a variety of tasks. The one on the right accepts a variety of programs, one program for each of a variety of tasks.

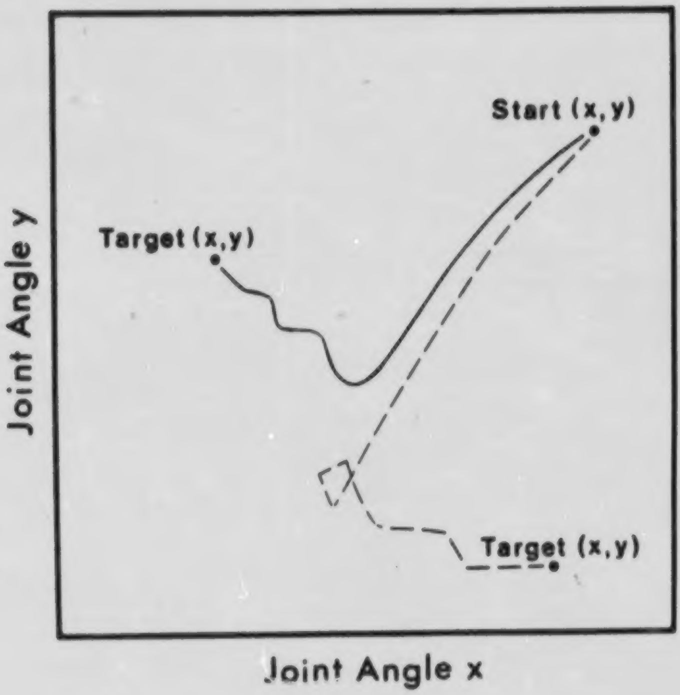

Start (x,y)
Target (x,y)
Target (x,y)
Joint Angle y
Joint Angle x

Figure 3: Exemplary solution strategies for two Krinskiy and Shik problems. The starting coordinates represent the angles of the subject's joints at the outset of the task and the target coordinates represent the values which minimize the function.

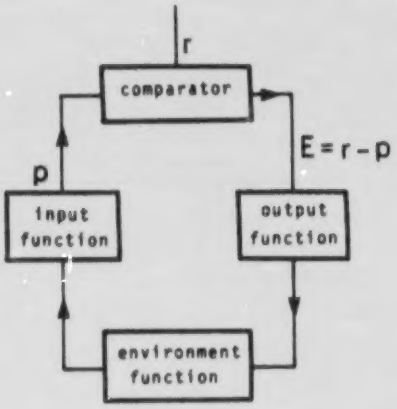

r
comparator
E = r - p
p
input function
output function
environment function

Figure 4: An individual control system.

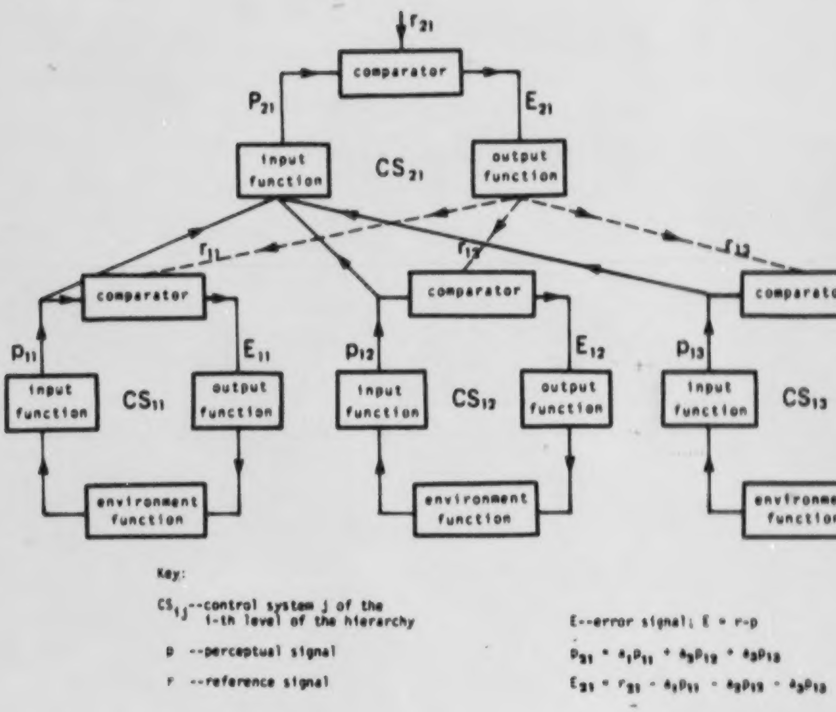

r_{21}
comparator
P_{21}
E_{21}
input function
CS_{21}
output function
r_{11}
r_{12}
r_{13}
comparator
comparator
comparator
p_{11}
E_{11}
p_{12}
E_{12}
p_{13}
E_{13}
input function
CS_{11}
output function
input function
CS_{12}
output function
input function
CS_{13}
output function
environment function
environment function
environment function
Key:
CS_{ij}--control system j of the i-th level of the hierarchy
p --perceptual signal
r --reference signal
E--error signal; $E = r-p$
$p_{21} = a_1p_{11} + a_2p_{12} + a_3p_{13}$
$E_{21} = r_{21} - a_1p_{11} - a_2p_{12} - a_3p_{13}$

Figure 5: A stack of control systems: three first-order systems nested under one second-order system.

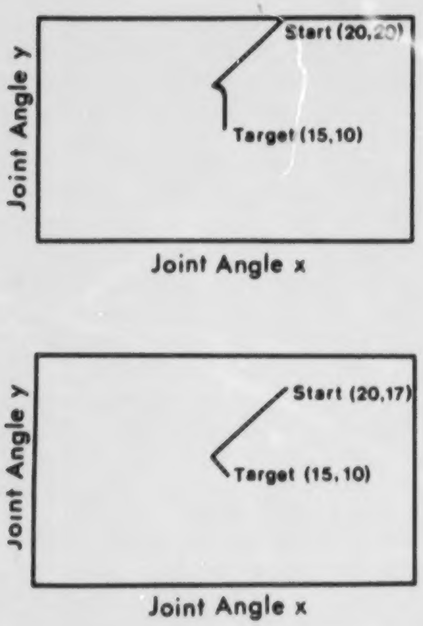


Figure 6: Movement strategies of the computer model (compare with Figure 3).

will be accompanied by relatively uncontrolled movements in the other passive links of the chain. Obviously, for a biokinematic chain such as an arm or a leg, muscular forces are not the only forces acting on the chain; besides gravity there are the kinetic energies and moments of force that necessarily accompany movements of the individual links.

A further and related principle of kinematic chains is that the design of a chain--the lengths and masses of its links, the manner of their joining and the degrees of freedom of the joints--determines the kind of curves that the chain can trace out over time. Now an actor can modify the design of a biokinematic chain and, therefore, its potential trajectories, in a very simple way: he can selectively freeze the degrees of freedom and vary the range of joint movement. The significance of this is that for any desired trajectory of a limb, elaborate control on the part of the actor--even moment-to-moment computation--may be needed to secure the trajectory given one "design" of the limb, yet very little computation may be needed given another, very different "design." The point is that, with an appropriate design, the reactive forces that are concomitant to movement of the chain as a whole may contribute significantly to the production of the trajectory, but with an inappropriate design, the reactive forces that accompany the chain's movement may contribute little to the desired trajectory and may even oppose it.

In this regard, consider the emergence of an effective sidearm strike pattern (hitting a ball baseball-style) in preschool children (see Wickstrom, 1970). The development of the skill is realized through the following changes: a more liberal swing due to an increase in the range of motion of the participating joints; increasing usage of the forward step or forward weight shift to initiate the strike pattern, and increasing pelvic and trunk rotation prior to the swing of the arms (in the earliest stages of acquisition, pelvic and trunk rotation occur as a result of the strike with the pattern being initiated by the arm motion). One way of looking at these changes is that they index transformations in the "design" of biokinematic chains. The two arms, coupled at the bat, constitute a biokinematic chain whose design is made more effective for the task by increased unlocking of the wrists and greater flexion at the elbows. The body as a whole is a biokinematic chain, the design of which is made more effective for striking by adding the degrees of freedom of trunk rotation and pelvic rotation. To paraphrase our remarks above, a more effective design of a limb or a body is one in which the reactive forces concomitant to movement are largely responsible for the achievement of the desired trajectory.

Another way of looking at these changes, however, observes that an actor, naive to a particular skill, curtails biokinematic degrees of freedom--through the complete immobilization of some joints (that are used when the skill is performed expertly) and a restriction on the range of motion of other joints--because he or she lacks a means of controlling the biokinematic degrees of freedom in the manner that the skill demands. It then follows that increasing expertise is indexed by a gradual raising of the ban on degrees of freedom (to borrow Bernstein's most apt phrase). Or, to put it slightly differently, increasing the number of controllable biokinematic degrees of freedom is synonymous with becoming more expert. As Bernstein (1967, p. 127) remarks: "The coordination of movement is the process of mastering redundant degrees of freedom of the moving organ, in other words its conversion to a controllable system."

In short, the changes indexing the acquisition of the batting skill can be interpreted in (at least) two ways: in one, as the converting of biokinematic degrees of freedom into controllable systems (coordinative structures), and in the other, as the designing of biokinematic chains so as to secure certain reactive forces. Surely these two interpretations are dual. By increasing the controllable degrees of freedom, the actor increases the potential variability of reactive forces that accompany the activity, thereby increasing the opportunity to discover what the activity-relevant reactive forces might afford by way of control. In the discovery of activity-relevant reactive forces, the actor prescribes the conversion of redundant degrees of freedom into controllable systems.

Let us summarize the tenor of these remarks. On the "Air Theory" formulation of coordinated activity, an executive must supply that control that the coordinative structures do not supply. On the "Ground Theory" formulation, an actor must supply that control that the external force field does not supply. In a blend of the two formulations we can say that, in the performance of an athletic skill, coordinative structures are so organized as to secure certain reactive forces; by the felicitous organization of coordinative structures the actor bends the force function that is given to yield the force function that is desired. In the grain-size of analysis prescribed by the event perspective, it is neither muscles nor joints that are coordinated in the performance of athletic skill, but forces--those supplied by the actor and those supplied by the environment.

ON CONVERTING BIOKINEMATIC FREE-VARIABLES INTO A CONTROLLABLE SYSTEM

In this third section we address the question of how an actor forms a controllable system [in Bernstein's (1967) terms] or a coordinative structure (in our terms). In the description of the actor developed in the second section, it was concluded that an act is more optimally described in terms of autonomous collectives of free-variables than in terms of the free-variables themselves, that is, the individual muscles or joints. To lay the groundwork for the analysis that follows, we identify three aspects of the problem of forming such collectives. These aspects are described abstractly; they are, however, reasonably intuitive. Moreover, they may be considered as fundamental aspects of all coordinative problems and we will attempt to show how they relate closely to the summary remarks of the first section.

Three Intuitions Relating to Action Problems

First, we believe that in a general but nontrivial sense, the problem of forming a coordinative structure or controllable system may be characterized, in part, in the following fashion: given an aggregate of relatively independent biokinematic degrees of freedom, how can the aggregate be so constrained, the individual degrees of freedom so harnessed, as to produce a particular, simple change in a particular single variable.[5] Thus, for example, to minimize the displacement of the point of intersection of the line of aim with the target, experienced marksmen constrain the joints of the weapon arm in such a fashion that the horizontal displacements of the individual kinematic links are reciprocally related (a form of constraint that is not at the disposal of the novice) (Arutyunyan, Gurfinkel and Mirskii, 1968, 1969). In paraphrase of the above intuition we may say, therefore, that the problem of forming a coordinative structure or controllable system is, in part, the problem of discovering the relevant constraint for a collection of many (fine-grain) variables, such as individual joints, that will realize a particular (coarse-grain) variable, such as a limb trajectory. In a somewhat different vernacular, it is the pattern of discovering the equivalence class of optimal combinations of these variables (Greene, 1969).

One easily appreciates that during the acquisition of a skill the fine-grain variables do not present themselves in precisely the same way every time. The specific details, that is, the initial conditions of the fine-grain variables, are not standardized. Nevertheless, the actor must select, on each occasion of the problem, one combination of the variables from the set of all possible combinations, and ideally, on each successive occasion the combination selected should approximate more closely the desired objective.

It is often remarked that the felicitous solution to problems of coordination is made possible by "knowledge of results" identified as information about whether an attempted solution (say, a particular movement) was right or wrong (qualitative knowledge of results), and if wrong, by how much (quantitative knowledge of results). Thus, Adams (1976) comments:

> The human learning of motor movement is based on knowledge of results, or information about error in responding. Knowledge of results can be coarse, like "Right" or "Wrong" or it can be fine grain, like "You moved 2.5 inches too long." (p. 216)

In our view this is a gratuitous claim. In the general case, information about degree of nearness to a desired outcome will be insufficient informational support for arriving at a solution to the coordination problem. Let us elaborate.

We identify the general case as discovering an optimal organization of, or constraint for, a number of free biokinematic variables. The argument can be made--consonant with the jigsaw puzzle metaphor--that for a system with η biokinematical degrees of freedom there ought to be at least η degrees of freedom in the information that supports the control of that system (Turvey et al., in press). These informational degrees of freedom can be most usefully understood as degrees of constraint (Turvey et al., in press). We can suppose, therefore, that discovering an optimal relation on η free-varying biokinematic degrees of freedom requires that at least n degrees of constraint be available perceptually. We may hypothesize that *in general, the ease and probability of discovering an optimal organization (that is, learning) relates directly to the extent to which degrees of constraint match degrees of freedom*.

---

[5] We owe this manner of describing controllable systems to H. H. Pattee (for example, 1970, 1973). He considers the existence of control constraints an essential and distinguishing property of living systems.

In discrete movement tasks (for example, Trowbridge and Cason, 1932), the actor must learn to move a limb or a limb segment a fixed distance. It is not difficult to imagine that in the acquisition of such simple tasks the actor freezes all the free-variables (joints) but one; that is, the actor manipulates a single biokinematic degree of freedom. The quantitative knowledge of results about how closely the movement approximated the desired distance is one degree of constraint that matches the one degree of freedom of the movement. Hence, in this case, quantitative knowledge of results is sufficient informational support for learning (see Adams, 1971). In the acquisition of an activity involving the regulation of more than one biokinematic degree of freedom, the single degree of constraint provided by quantitative knowledge of results would be inadequate. The fundamental point is this: quantitative knowledge of results specifies, in a limited sense, what *not* to do next but, significantly, it does not specify *what* to do next. The novice golfer who putts two meters to the right of the hole sees that he has erred, but this information, *in and of itself*, cannot tell him *how* to change the organization of his biokinematic free-variables so as to err less on the next occasion. If quantitative knowledge results were the only source of constraint on selecting combinations of biokinematic free-variables, then we may suppose that the search for the optimum combination would be essentially blind (that is, the combinations would be chosen at random) and, in principle, the search could proceed indefinitely.

A remedy for the inadequacy of quantitative knowledge of results is suggested by the two remaining notions. On the acceptance of the actor as a special-purpose problem solver, Gel'fand and Tsetlin (1962, 1972) asked what it is that might characterize, in general, the problems posed to the actor so that he might bring to bear *specialized* search procedures, tailor-made (presumably in the course of evolution) for such problems. They suggest that the actor might operate on the tacit assumption that the problems he encounters are *well-organized* in the sense that (1) the variables indigenous to a problem may be partitioned into essential (intensive) and nonessential (extensive) variables, and (2) that a variable is consistently a member of one or the other class. Given the assumption that the problem is well-organized, the actor can successfully apply a certain method of search through the space of constraints (for Gel'fand and Tsetlin it is the Ravine Method that is described below). The actor initiates the specialized search method ignorant of the actual pattern of organization of the problem; it is only in the course of the search that the pattern is disclosed (Gel'fand and Tsetlin, 1962).

Our second intuition, therefore, is that in a general but nontrivial sense, each and every problem confronting the skill-acquirer may be characterized as follows: *with reference to the objective*, there is an organization defined on the participating elements. The organization may be described as a function that is preserved invariantly over changes in the specific value of its variables. We will speak, therefore, of the organizational invariant of a coordination problem. An invariant may be usefully defined for our purposes as information about something, in the sense of specificity to that something, that is preserved over relevant transformations (see Gibson, 1966; Shaw, McIntyre and Mace, 1974). By implication, the style of change imposed by an actor on the aggregate of variables is significant to the determination (detection) of the organizational invariant; put bluntly, not all classes of change will reveal the organizational invariant (see footnote 7).

The third intuition relates to the issue of how a search through combinations of many variables may be guided. Whatever we imagine the search method to be, it must necessarily be the case that the successive "experiments" conducted on the variables exploit information realized by the experiments. Our third intuition, therefore, is that in a general but nontrivial sense, there is available to the actor seeking to solve a coordination problem, information that specifies, relatively precisely, what to do next. Such information, we believe, may often take the form of abstract relations defined over variables of stimulation over time, and that becoming attuned to such information is part of the solution--developing, pari passu, with the isolating of the organizational invariant.

Let us relate the above three essential components of the acquisition of a controllable system to the concluding remarks of the first section, as follows:

1) An actor learns to make of himself a "special-purpose device" designed optimally for the task at hand. He does so by discovering an appropriate organization of his musculature that differs for different acts (for example, walking versus swimming).

Several sets of muscle-organizations may suffice to get a given job done, but some may be more efficient than others. For example, an actor learns to swim before he learns to swim skillfully. Following the work of Gel'fand and Tsetlin cited earlier, we suppose that species have evolved special strategies for selecting the most harmonious organization of muscle-systems among the restricted set of possible ones. Thus, the idea of the actor as a special-purpose device applies not only to the individual actor acquiring a particular skill, it also applies to the class of actors acquiring any skilled act. At this more coarse-grained level of description, any problem of skilled action may be described in part as a problem of optimizing a function of several variables (see above).

2) The skill to be acquired may be described as a set of potential constraints on the character of an event (as an organizational invariant). These constraints set boundary conditions on the possible muscle organizations that the actor can invoke to achieve his performance aims. Therefore, the actor's discovery of the organizational regularities of a task vastly simplifies his search for an optimal self-organization.

3) The efforts of a novice to perform an act may be viewed, in part, as discovery or search tactics aimed at revealing the organizational structure of the task.

The Experimental Task

The task that we have been investigating was designed by Krinskiy and Shik (1964). A subject is seated before a scale and instructed to make the scale-indicator point to zero. He controls the indicator position in this way: two of his joint angles (typically his elbow joints) are monitored continuously. The values of the two angles are input to a computer that transforms them according to the mapping: $E = |x-y-(a-b)| + \alpha|x-a| + \alpha|y-b|$; x and y are variables that take on the values of the joint angles each time they are sampled; a, b, and α are parameters that are changed across, but not within, experiments or trials. The equation controls the needle position on the scale. That is, the needle position corresponds in some simple way to E. The subject can make the needle on the scale go to zero by finding the angles of his joints for which the mapping takes on the value $E = 0$. The needle points to zero when the subject has minimized the mapping.

The subject is unaware of the specific nature of the control that he has over the needle. He knows that by changing his joint angles he changes the needle position. However, he does not know that his joint angles are the x and y coordinates of some mapping whose output corresponds to the position of the scale-indicator. The starting position of the subject's joint angles may be varied or kept the same over trials. Likewise, the target values (the values of his joint angles at which the function is minimized) may be varied or maintained over trials.

Krinskiy and Shik provide a limited quantity of data in the form of graphs that depict the solution strategies of their subjects. Sample graphs are shown in Figure 3. The x-axis represents the value of one elbow-joint angle and the y-axis the value of the other. A diagonal line on the graph represents simultaneous changes of the joint angles on the part of the subject, while horizontal or vertical lines represent a change in just one angle. (The slopes of the lines in Figure 3 indicate that the rates of change of the two joint angles are the same; the slopes are approximately equal to one.) As the subjects approach the solution, they begin changing the values of the two joint angles individually.

Although the minimization task may seem an artificial one, it does have the essential components of a problem of skill acquisition that we have outlined. First, the equal velocities of the movement of the two forearms suggest an organization of the subjects' musculature that spans both joints (see Kots and Syrovegin, 1966). In addition, an attractive property of the task for the purposes of investigation is that its organizational invariant is known to the investigator. (It is the mapping $E = |x-y-(a-b)| + \alpha|x-a| + \alpha|y-b|$.) However, it is not known to the subject until his own movements reveal it to him as a lawful, though complex relationship between the changes of his joint angles and the movement of the needle on the scale. Apparently when the actor has learned the task, he controls the performance of a muscle-system. We will suggest that he does so by detecting the higher-order properties of optical stimulation that prescribe what he should do next, given his aim to set the scale-indicator to zero.

A final attractive property of the task is that it engages the subject in a search for the minimum of a function of several variables. In this regard it mimics a task that Gel'fand and Tsetlin (1962, 1971; see also Gel'fand, Gurfinkel, Tsetlin and Shik, 1971) argue is characteristic of muscle-systems as they seek a maximally harmonious self-organization. An organization of muscle-systems that is maximally harmonious may be one in which the activities governed by the different muscle-systems do not compete. If we represent the interactions among the muscle systems as variables, then the search for a harmonious self-organization may be conceptualized as the search for the minimum of a function that encompasses the variables. Gel'fand and Tsetlin suggest that a set of search tactics has evolved, which they call ravine tactics, that are tailored to this kind of optimization task, although they may be ill-suited to other ones. We will describe these search tactics shortly. Here we merely note that the task of Krinskiy and Shik may not be, in fact, an artificial one in which an actor will engage. Indeed, it was devised to assess whether or not actors employ ravine tactics when given a task for which the tactics are especially suited.

Our contribution to the investigation of the minimization task has been to ask how a subject might learn to solve it efficiently. We have done so by modeling, with the aid of a computer, a skilled performer of the task. Instead of modeling directly the superficial properties of the strategy depicted in Figure 3, we attempted more simply to design a model that could perform the task without invoking blind or random search tactics. Our model uses a strategy that in its superficial properties is similar to the one depicted in Figure 3. The model initially changes both angles at a constant equal rate and as it nears the target values, changes the angles individually. It adopts this way of doing the task as a by-product of a deeper strategy--which is to exploit the higher-order variables of optical stimulation offered by the changes in the scale-indicator over time, in preference to the relatively uninformative value E given by the instantaneous needle position.

Before looking at this model, it is instructive to look at one that evidently cannot perform the task without invoking random search tactics (hence the model never becomes a skilled performer). This latter model is of interest because it is the model of Powers (1973) and it is consistent with the models of closed-loop motor performance proposed, for instance, by Adams (1971) and described by Greenwald (1970).

By showing that a model consistent with these theories cannot solve the task in a plausible way, we do not mean to imply that actors never use quantitative knowledge of results (here the value E) to regulate their motor performances. Indeed, the evidence cited by Adams (1971) and by Greenwald (1970) suggest this as a potent source of information in the acquisition of some skilled movements. We only wish to propose that actors are flexible and can adapt their acquisition strategies, within limits, to the useful dimensions of information provided by a particular problem.

We have selected the model of an actor/perceiver developed by Powers (1973) to serve as a prototypical model of closed-loop motor performance. This and other models of closed-loop performance evidently are general-purpose devices by virtue of having a single general-purpose acquisition strategy. We will show that the strategy is inappropriate to the solution of the task devised by Krinskiy and Shik; and we will suggest that its inapplicability extends to any skilled performance in which higher-order variables of stimulation provide the useful and controlling dimensions of information to an actor.

A Model of Closed-Loop Motor Control: Powers, 1973

For Powers, the nervous system of an actor/perceiver may be characterized as a hierarchy of control systems. Figure 4 depicts the structure of an individual control system. Each system works to realize a particular *perceptual* state of affairs and it accomplishes its aim in the way that a mechanical homeostatic device does. Its intent (its intended perceptual state of affairs) constitutes a reference signal, r, for the system. That signal is compared periodically with the actual perceptual state of affairs, p. Both sources of information to the control system, the reference signal and the perceptual signal, are conceptualized as quantities, in particular, as rates of neural firing.

The two quantities, r and p, are subtracted in a comparator and the difference constitutes an error signal, E. If the value of E is nonzero, it is transformed into an output signal, or correction procedure, that effects changes in the environment of the control system. (E constitutes the address in memory of a stored correction procedure.) In turn, the environmental changes alter the perceptual input to the control system in the direction of the reference signal. If the actual and intended perceptual states of affairs are the same, $E = 0$, and the control system has achieved its intent.

A condition for the successful performance of the model is that an error signal must correspond in a one-to-one, or, in a nearly one-to-one, way with an appropriate correction procedure. That is, an error signal must specify what needs to be done to nullify it. Apparently this condition is met in the positioning tasks investigated by Adams (1971) and in the line drawing tasks of Trowbridge and Cason (1932). In these tasks, when the experimenter provides quantitative knowledge of results, the subject is given information that specifies what he must do to rectify his error. Similarly, in the tracking tasks described by Powers (1973), the perceived difference in locations of a target spot of light and a cursor specify what must be done to close the gap.

However, the condition is not met in the minimization task of Krinskiy and Shik. In that experiment, the error signal E does not specify to the subject what he must do to correct it. To take just one example, consider the values of E when a, b, and α, the parameters of the mapping, are set to 15, 10, and .2, respectively. The mapping is minimized when $x = a = 15$, and $y = b = 10$. Table 1 displays a set of values of x and y for which the error signal is invariantly 6. In the first case, the joint angle corresponding to the value of y is at its target position. In order for the joint angle corresponding to the variable x to reach its target position of 15, x has to be increased in value by 5. Hence, the correction procedure that is stored in a memory location whose address is $E = 6$, should specify no change in the variable y and an increase of 5 units in the value of x. That correction procedure is inappropriate to all of the other cases listed in Table 1. To correct an error of 6 when $x = 12$ and $y = 12$, for instance, x has to be increased in value by 3 and y decreased by 2. To correct an error of 6 when $x = 15$ and $y = 15$, x has to remain unaltered and y has to be decreased in value by 5. To correct an error of 6 when $x = 18$ and $y = 8$, x has to be *decreased* by 3 and y increased by 2. Finally, when $x = 30$ and $y = 25$, both have to be decreased in value by 15. These examples do not exhaust the ways in which an error of six can be obtained, nor is six the only ambiguous error signal.

TABLE 1: Some ways of obtaining an error of 6 in the mapping:
$E = |x-y-(15-10)| + .2\,|x-15| + .2\,|y-10|$

Component of the mapping			Correction procedure	
x	y	E	x	y
10	10	6	x = x + 5	
11	11	6	x = x + 4	y = y - 1
12	12	6	x = x + 3	y = y - 2
13	13	6	x = x + 2	y = y - 3
14	14	6	x = x + 1	y = y - 4
15	15	6		y = y - 5
18	8	6	x = x - 3	y = y + 2
17	7	6	x = x - 2	y = y + 3
16	6	6	x = x - 1	y = y + 4
15	5	6		y = y + 5
30	25	6	x = x -15	y = y -15

In short, for the cases presented in Table 1, different correction procedures appropriately correspond to the same error signal of 6. The quantitative knowledge of results that the error signal provides gives little or no information about how it can be nullified, and hence knowledge of results in this task is of limited utility to a subject. On the other hand, ΔE or the velocity of the moving needle on the scale does provide useful information to a subject, as we will show. However, ΔE information is provided only over successive movements of the actor and over successive loops around the control system, and the individual control system of the kind that Powers describes uses only the current value of E to guide its behavior.

Power's model and other closed-loop models appear to exclude the use of higher-order relationships that are revealed over relatively long stretches of time between the movements of the actor and their optical or other perceptual concomitants. Furthermore, we can show that the ambiguity and uninformativeness of quantitative knowledge of results is not peculiar to the task of Krinskiy and Shik. Rather, it is general to most complex tasks, particularly if they are considered to be performed by a hierarchy of closed-loop systems.

Quantitative Knowledge of Results is Equivocal in Hierarchical Closed-Loop Systems

In the model of Powers, the nervous system is a nested set of control systems of which only the lowest-level (first-order) systems are in direct contact with the environment. The first-order systems extract information about intensity of stimulation at the receptors. More abstract properties of stimulation (for instance, its form or temporal properties) are constructed by the second- to ninth-order systems based on the first-order perceptual signals. Each superordinate system receives input from several systems on the next level down. It combines them according to some linear transformation that is peculiar to it. The outcome of the linear transformation is a higher-order property of the stimulus input than had been extracted by any of the subordinate systems.[6]

At every level of the system, perceptual signals are subtracted from reference signals, the latter representing an intended perceptual state of affairs. The resulting error signal constitutes the address of a stored correction procedure. For a first-order system, the correction procedure effects real changes in the environment of the actor. The correction procedures of the higher-order control systems constitute reference signals for lower-order systems. That is, higher-order systems effect changes in the world only indirectly by changing the reference signals of lower-order systems.

It is easy to show that error signals must almost invariably be ambiguous with respect to their appropriate correction procedures in a hierarchical model of this sort. Figure 5 demonstrates this with a two-tiered nervous system.

Consider a nervous system composed of three first-order systems (CS_{11}, CS_{12}, CS_{13}) and one second-order system (CS_{21}). Each first-order system supplies CS_{21} with a perceptual signal. According to the model, the perceptual signal of the second-order system, p_{21}, is a linear transformation of the three, first-order perceptual signals, p_{11}, p_{12}, p_{13}. Thus, $p_{21} = a_1p_{11} + a_2p_{12} + a_3p_{13}$. That signal is subtracted from the reference signal, r_{21}, of the second-order system. The result, $E = r_{21} - a_1p_{11} - a_2p_{12} - a_3p_{12}$, is the error signal of the second-order system. It constitutes the address of a stored correction procedure that will provide the reference signals, r_{11}, r_{12} and r_{13}, of the first-order systems.

---

[6] Powers' claim is not unlike that of feature-based theories of visual perception. It is that the abstract, higher-order properties of the world are constructed (rather than being detected) by perceptual systems. The raw material for the constructions are lower-order, primitive properties of the world that perceptual systems detect directly. This claim is in contrast to that of Gibson (1966) and others (for example, Turvey, 1977a). Gibson holds that any properties of a world that an organism *perceives*, however abstract they may be, are detected by it directly.

We should point out an apparent flaw in Powers' and the feature-based views. Consider a perceptual system that has detected n primitive elements and that is now given the task of constructing a higher-order percept from them.

Even if the domain of possible combinations of the n primitive elements is confined to those in two-space and to ordinal relationships among them, there are n! possible organizations of the elements. If we expand the domain to include the third spatial dimension and if we assign significance to the distances among elements, the number of possible organizations of the n primitives must escalate dramatically. Powers' theory has to endow an organism with the means of selecting the single *actual* organization of the elements out of the potentially astronomical number of possibilities. A theory can avoid endowing an organism with this mystical ability if it recognizes that the sector of the world being observed gives these hypothetical primitives only one organization. A plausible proposal is that the observer *detects* the abstract properties themselves, rather than having to build them out of a number of primitives.

For concreteness, consider an error signal, $E = 6$. There are very many possible combinations of values for r_{21}, p_{11}, p_{12} and p_{13} that might yield a value of six, even if some boundaries are set on the possible ranges of values that each might take. The error signal might be entirely due to an error of one of the first-order systems; or it could be due to various combinations of pairs of first-order systems; or it could be one of many combinations of errors on the part of all three first-order systems.

Quantitative knowledge of results must rarely be informative in a hierarchical closed-loop system because, typically, there is a one-to-many mapping between an error signal and the conditions that may have provoked it. We can conclude from that, perhaps, that the actor/perceiver is not appropriately characterized as a hierarchy of control systems, at least when he is performing tasks in which he must exploit the abstract information putatively extracted by the superordinate levels of the system.

The closed loop model of Powers characterizes the actor as an inflexible general purpose device. Let us turn now to a different type of model that purports to govern only a limited class of activities. Its performance strategy is tailored to the special features of that limited class of acts but is inappropriate to activities outside of that class. The model that performs the minimization task characterizes just one among the many special-purpose devices that an actor can become, depending on his performance aims.[7]

Searching the Two-Variable Space: The Ravine Method

For Gel'fand and Tsetlin (1962, 1971), a strategy that is tailored to the minimization problems of muscle systems is the Ravide Method. It combines local and nonlocal search tactics and thereby avoids the tendency of strictly local search methods to be deceived by local minima of a search space.

The method works in the following way. A local search strategy is selected. (In the problem of Krinskiy and Shik, the actor selects some way of altering the values of his joint angles.) The strategy is maintained until the value $\Delta E/E$ reaches some preselected lower bounds. A small value of $\Delta E/E$ implies that the current strategy has reached a point of diminishing returns.

---

[7] We should point out that the current model is of a skilled performer of the task. An aim of our preliminary efforts has been to characterize the state towards which a novice is working. By establishing the ways in which a skilled performer coordinates the movements of his limbs in relation to the variables of stimulation provided him by the scale-indicator, we can specify the variables of stimulation to which the novice must become sensitive if he is to learn to perform the task skillfully. Clearly, the discovery tactics of the novice must be such that they reveal that organizational invariant (that is, the invariant relationship between what he does and what he sees).

subordinate systems.[6]

At every level of the system, perceptual signals are subtracted from reference signals, the latter representing an intended perceptual state of affairs. The resulting error signal constitutes the address of a stored correction procedure. For a first-order system, the correction procedure effects real changes in the environment of the actor. The correction procedures of the higher-order control systems constitute reference signals for lower-order systems. That is, higher-order systems effect changes in the world only indirectly by changing the reference signals of lower-order systems.

It is easy to show that error signals must almost invariably be ambiguous with respect to their appropriate correction procedures in a hierarchical model of this sort. Figure 5 demonstrates this with a two-tiered nervous system.

Consider a nervous system composed of three first-order systems (CS_{11}, CS_{12}, CS_{13}) and one second-order system (CS_{21}). Each first-order system supplies CS_{21} with a perceptual signal. According to the model, the perceptual signal of the second-order system, p_{21}, is a linear transformation of the three, first-order perceptual signals, p_{11}, p_{12}, p_{13}. Thus, $p_{21} = a_1p_{11} + a_2p_{12} + a_3p_{13}$. That signal is subtracted from the reference signal, r_{21}, of the second-order system. The result, $E = r_{21} - a_1p_{11} - a_2p_{12} - a_3p_{12}$, is the error signal of the second-order system. It constitutes the address of a stored correction procedure that will provide the reference signals, r_{11}, r_{12} and r_{13}, of the first-order systems.

---

[6] Powers' claim is not unlike that of feature-based theories of visual perception. It is that the abstract, higher-order properties of the world are constructed (rather than being detected) by perceptual systems. The raw material for the constructions are lower-order, primitive properties of the world that perceptual systems detect directly. This claim is in contrast to that of Gibson (1966) and others (for example, Turvey, 1977a). Gibson holds that any properties of a world that an organism perceives, however abstract they may be, are detected by it directly.

We should point out an apparent flaw in Powers' and the feature-based views. Consider a perceptual system that has detected n primitive elements and that is now given the task of constructing a higher-order percept from them.

Even if the domain of possible combinations of the n primitive elements is confined to those in two-space and to ordinal relationships among them, there are n! possible organizations of the elements. If we expand the domain to include the third spatial dimension and if we assign significance to the distances among elements, the number of possible organizations of the n primitives must escalate dramatically. Powers' theory has to endow an organism with the means of selecting the single actual organization of the elements out of the potentially astronomical number of possibilities. A theory can avoid endowing an organism with this mystical ability if it recognizes that the sector of the world being observed gives these hypothetical primitives only one organization. A plausible proposal is that the observer detects the abstract properties themselves, rather than having to build them out of a number of primitives.

For concreteness, consider an error signal, $E = 6$. There are very many possible combinations of values for r_{21}, p_{11}, p_{12} and p_{13} that might yield a value of six, even if some boundaries are set on the possible ranges of values that each might take. The error signal might be entirely due to an error of one of the first-order systems; or it could be due to various combinations of pairs of first-order systems; or it could be one of many combinations of errors on the part of all three first-order systems.

Quantitative knowledge of results must rarely be informative in a hierarchical closed-loop system because, typically, there is a one-to-many mapping between an error signal and the conditions that may have provoked it. We can conclude from that, perhaps, that the actor/perceiver is not appropriately characterized as a hierarchy of control systems, at least when he is performing tasks in which he must exploit the abstract information putatively extracted by the superordinate levels of the system.

The closed loop model of Powers characterizes the actor as an inflexible general purpose device. Let us turn now to a different type of model that purports to govern only a limited class of activities. Its performance strategy is tailored to the special features of that limited class of acts but is inappropriate to activities outside of that class. The model that performs the minimization task characterizes just one among the many special-purpose devices that an actor can become, depending on his performance aims.[7]

Searching the Two-Variable Space: The Ravine Method

For Gel'fand and Tsetlin (1962, 1971), a strategy that is tailored to the minimization problems of muscle systems is the Ravide Method. It combines local and nonlocal search tactics and thereby avoids the tendency of strictly local search methods to be deceived by local minima of a search space.

The method works in the following way. A local search strategy is selected. (In the problem of Krinskiy and Shik, the actor selects some way of altering the values of his joint angles.) The strategy is maintained until the value $\Delta E/E$ reaches some preselected lower bounds. A small value of $\Delta E/E$ implies that the current strategy has reached a point of diminishing returns.

---

[7] We should point out that the current model is of a skilled performer of the task. An aim of our preliminary efforts has been to characterize the state towards which a novice is working. By establishing the ways in which a skilled performer coordinates the movements of his limbs in relation to the variables of stimulation provided him by the scale-indicator, we can specify the variables of stimulation to which the novice must become sensitive if he is to learn to perform the task skillfully. Clearly, the discovery tactics of the novice must be such that they reveal that organizational invariant (that is, the invariant relationship between what he does and what he sees).

When the criterial $\Delta E/E$ is attained for the first time, the actor alters his strategy randomly. The new strategy is maintained until $\Delta E/E$ again reaches its criterial value. The next strategy shift (ravine step) is selected, based on which of the previous two was the more successful in approximating the function's minimum. The ravine step is taken in a direction that is nearer to whichever of the first two strategies was the more successful. The procedure is continued until the minimum is reached.

This optimization procedure exploits the special properties of those multi-variable functions that, according to Gel'fand and Tsetlin, characterize the muscle-systems of an actor. (One special property is that the mapping is "well-organized" in the sense described above.) The form of the search methods used by the subjects of Krinskiy and Shik and depicted in Figure 3 is compatible with the hypothesis that they use ravine tactics.

A similar search procedure is also compatible with the graphs in Figure 3. We devised this latter procedure initially as a way of translating the principles of the ravine method, expressed as a set of computational procedures by Gel'fand and Tsetlin, into a set of principles of joint-angle movement that could be implemented by an actor. In doing so, we discovered information provided by the scale-indicator that may be more useful to a performer of the task than is $\Delta E/E$. The final model that we will describe rarely shifts its search strategy blindly, as that of Gel'fand and Tsetlin does on the first ravine step. It avoids having to do so by maximally exploiting the information provided by the values ΔE and $\Delta(\Delta E)$, the velocity and the acceleration of the scale-indicator. These properties of the event in which the performer participates prescribe to him what he should do next, given his performance aims.

Searching the Two-Variable Space: Sensitivity to Rate of Change and Rate of Rate of Change

The model is instantiated as a computer program that has available eight possible strategies of joint-angle movement. Four strategies change both joint angles simultaneously and the other four change just one of the angles. The four strategies of simultaneous movement are to increment both angles, decrement both, increment the angle corresponding to the variable x and to decrement y, and to decrement x and increment y. The four strategies of the second type are to increment or decrement x or y.

On each pass, the program alters the value of x and/or of y in the direction dictated by its current muscular organization--that is, by its choice of movement strategy. The two joints are potentially a single coordinative structure; hence, it is simplest for the model/performer to move his two forearms at the same rate.

After altering the values of x and y by equivalent amounts on each pass, the new value of E is computed. If $E = 0$, the program halts because the function has been minimized. If E is nonzero, the values of ΔE (the current value of E subtracted from its previous value) and of $\Delta(\Delta E)$ (the currect value of E subtracted from its previous value) are computed. These higher order properties of the moving scale-indicator provide a fairly rich source of information to the model that uses it to guide its next step.

The Information Provided by ΔE

Let us look separately at the three components of the organizational invariant $E = |x-y-(a-b)| + \alpha|x-a| + \alpha|y-b|$, as a way of seeing how the various higher-order variables of stimulation specify to a skilled performer what he is to do next. The first component of the mapping, $C_1 = |x-y-(a-b)|$, is least useful because its contribution to the movement of the needle on the scale provides primarily "proprioceptive" information and little information about whether the movement strategy is working or not. In contrast, $C_2 = \alpha|x-a|$ and $C_3 = \alpha|y-b|$ provides "exteroceptive" information; the coefficients, α, of C_2 and C_3 are different from that of C_1, and therefore their contribution to the value of ΔE can be distinguished from C_1's contribution.

Table 2 provides some examples of the information provided by E. Six cases are represented in the table. Three correspond to a movement strategy in which the performer increments the values of both joint angles, and three correspond to a strategy in which he increments x and decrements y. The remaining strategies of simultaneous movement may be observed by reading the cases from bottom to top. For each strategy, in one instance represented in the table, the strategy is correct for both joint angles (Ia and IIa in Table 2). That is, both angles are approaching their target values. (In the examples given, the target values are $x = 15$ and $y = 10$.) In a second instance (Ib and IIb), the strategy is appropriate for x, but not for y, and in the last instance (Ic and IIC), it is appropriate for neither. The first component of the mapping, $C_1 = |x-y-(a-b)|$, contributes a value of 0 to the scale-indicator velocity (ΔE), if both joint angles are incrementing or if both are decrementing (I a-c in Table 2). It contributes a value of 2 (more generally, twice the value of the coefficient of C_1) if one angle is being incremented and one decremented (II a-c). Thus, the proprioceptive information that the subject obtains from the scale-indicator tells him whether or not his joint angles are moving in parallel. The sign of the contribution of C_1 to ΔE (that is, the direction of needle movement) provides general information about whether or not the current strategy is working to make the needle point to zero.

The other two components of the mapping, $C_2 = \alpha|x-a|$, and $C_3\ \alpha|y-b|$, contribute exterospecific information to the value E. Independently of the particular movement strategy that the performer has adopted, they contribute values to ΔE that are different, depending on whether both joint angles, just one joint angle, or neither joint is approaching the target. If both angles are approaching their targets (Ia and IIa in the table), C_2 and C_3 contribute a value of -2α. When one angle is moving towards its target (Ib and IIb), then the contribution of C_2 and C_3 is zero, because one contributes α and the other $-\alpha$. The contributions of C_2 and C_3 can be distinguished from that of C_1 because their coefficients are different from C_1's coefficient (here $\alpha = .2$).

Let us briefly consider an example that illustrates how the value of ΔE can guide the movements of a skilled performer of the task. If the value of ΔE is 2, the skilled performer knows two things. First, he knows that his joint angles are changing in opposite directions (one is incrementing and one is decrementing). In addition, because C_2 and C_3 are not represented in the needle velocity, he knows that only one of his joint angles is moving towards its target value. He then should alter the direction of movement of just one

TABLE 2: Information provided by scale-indicator velocity, (ΔE).

Pass through the computer program	Component of the mapping							
	X	Y	C_1	C_2	C_3	E	ΔE	(ΔE)
	I.a. Incrementing X, Y: appropriate strategy							
1	10	5	0	1	1	2		
							.4	
2	11	6	0	.8	.8	1.6		0
							.4	
3	12	7	0	.6	.6	1.2		
	I.b. Incrementing X, Y: appropriate only for X							
1	10	11	6	1	.2	7.2		
							0	
2	11	12	6	.8	.4	7.2		0
							0	
3	12	13	6	.6	.6	7.2		
	I.c. Incrementing X, Y: inappropriate strategy							
1	16	12	1	.2	.4	1.6		
							-.4	
2	17	13	1	.4	.6	2.0		0
							-.4	
3	18	14	1	.6	.8	2.4		
	II.a. Incrementing X, Decrementing Y: appropriate strategy							
1	10	14	9	1	.8	10.8		
							2.4	
2	11	13	7	.8	.6	8.4		0
							2.4	
3	12	12	5	.6	.4	6.4		
	II.b. Incrementing X, Decrementing Y: appropriate only for X							
1	10	4	1	1	1.2	3.2		
							2	
2	11	3	3	.8	1.4	5.2		0
							2	
3	12	2	5	.6	1.6	7.2		
	II.c. Incrementing X, Decrementing Y: inappropriate strategy							
1	17	5	7	.4	1	8.4		
							-2.4	
2	18	4	9	.6	1.2	10.8		0
							-2.4	
3	19	3	11	.8	1.4	13.2		

angle. It cannot be determined which one should be altered because the coefficients of C_2 and C_3 are the same. If the performer happens to choose the correct angle to change, on the next pass ΔE will equal .4 (that is, 2α), indicating that the angles are now changing in parallel and that both are moving toward their targets. If the choice was incorrect, $\Delta E = -.4$, and the performer knows to shift the direction of movement of both angles.

The Contribution of (ΔE)

The velocity of needle movement changes when one of the actor's joint angles reaches and goes beyond its target value. Consider the example in Table 3. In the example, both joint angles are being incremented. Hence C_1 contributes a value of zero to ΔE. In addition, on going from the first pass to the second, both angles are approaching their target values; hence C_2 and C_3 contribute a value of .4 to ΔE. Going from the second pass to the third, however, x moves away from its target value of 15, while y continues to approach its target value of 10. Therefore C_2 contributes -.2 and C_3, + .2 to the value of ΔE. The new ΔE is zero, and $\Delta(\Delta E)$ is .4. This deceleration of the needle is an indication that one of the two angles has reached and surpassed its target. When that occurs, the model shifts from a strategy of simultaneous movement of both joint angles to one of changing a single joint angle.[8]

TABLE 3: Information provided by scale-indicator deceleration $\Delta(\Delta E)$.

Pass through the computer program	Component of the mapping				
	X	Y	E	ΔE	(ΔE)
1	14	6	4		
				.4	
2	15	7	3.6		.4
				0	
3	16	8	3.6		

Figure 6 displays the movement strategies of our model. They are similar in form (but are more efficient than) those of the subjects of Krinskiy and Shik depicted in our Figure 3.

---

[8] The human subjects in the experiments of Krinskiy and Shik typically shifted from a strategy of simultaneous change of the two joint angles to one of successive change as they neared the target. The strategy of simultaneous change best reveals the "organizational invariant" of the task, and therefore is an optimal strategy of movement until one target value is reached.

CONCLUDING REMARKS

Our model and the mathematical model of Gel'fand and Tsetlin both perform the minimization task successfully. Furthermore, in their superficial properties, the strategies of these two models match the performance of the subjects of Krinskiy and Shik. Both models perform the minimization task by adopting procedures that are tailored to the special features of that task, but that are inappropriate to the features of other ones. We can perhaps conclude from these observations, and from the apparent incapacity of Powers' model to solve the task in an efficient way, that the human subjects in the experiment of Krinskiy and Shik likewise adopted a task-specific strategy. Given that those human subjects presumably are capable of performing other kinds of acts for which these tactics must be inappropriate (for example, the positioning task of Adams, 1971), we may consider this work to provide preliminary support for our conception of the actor as a general-purpose device by virtue of the capacity to become a variety of special-purpose devices.

The procedures of our model are distinguished from the mathematical optimization procedures of Gel'fand and Tsetlin in a way that seems significant to us. We suggested a principle (see also Turvey et al., in press) which holds that for the degrees of freedom necessitating control, there must be at least as many degrees of constraint in the information supporting that control. We suggested also that the two sources of control constraints are the environment and the actor (second section).

In the model of Gel'fand and Tsetlin, as applied to the minimization task of Krinskiy and Shik, degrees of constraint are largely supplied by the actor. The environment supplies the values ΔE and E, whose ratio guides the actor's selection of a new strategy. However, its guidance is minimal. That is, the ratio $\Delta E/E$ tells the actor when he should adopt a new strategy of movement, but it does not prescribe which strategy he should select. The actor selects a ravine step based on calculations on his part that compare the degrees of success of the two preceding sets of local search tactics.

Relative to this minimal use of environmental sources of constraint, our model yields up more of the responsibility for control. The environmentally-given values ΔE and $\Delta(\Delta E)$ not only tell the actor when to shift strategies, they also prescribe how he should alter his strategy to achieve his aim. In short, in this model, relative to that of Gel'fand and Tsetlin, the actor supplies few degrees of constraint and the environment supplies correspondingly many.

We find it intriguing to speculate that these two models may characterize actors at different phases of the skill-acquisition process. The model of Gel'fand and Tsetlin may characterize an actor who is sufficiently skilled to solve the task, but who does not yet perform it in the most efficient way. The actor provides some degrees of constraint that the environment would provide were he organized or attuned to detect them. Our model, in contrast, yields up to the environment as much of the responsibility for control as we have been able to uncover.

In the second section of the present paper, we sought to outline the kinds of information available to an actor, given an optimal level of description of the environmentally structured energy distributions that surround him. The potential sources of controlling information available to an actor in a natural environment exceed in number and in level of abstraction those sources made available to the actor in the experiment of Krinskiy and Shik. Nevertheless, as we have shown, the relatively limited information manifest in the Krinskiy and Shik task can tightly constrain the performance of that task. Collectively, these concluding remarks reiterate a major theme of the present paper, namely, that a careful examination of the environment as a perceptually specified source of constraint is mandatory to the understanding of the acquisition and performance of skilled activity.

REFERENCES

Adams, J. (1971) A closed-loop theory of motor learning. *Journal of Motor Behavior 3*, 111-149.

Adams, J. (1976) Issues for a closed-loop theory of motor learning. In *Motor Control: Issues and Trends*, ed. by G. Stelmach. (New York: Academic Press).

Arutyunyan, G. A., V. S. Gurfinkel and M. L. Mirskii. (1968) Investigation of aiming at a target. *Biophysics 13*, 642-645.

Arutyunyan, G. A., V. S. Gurfinkel and M. L. Mirskii. (1969) Organization of movements on execution by man of an exact postural task. *Biophysics 14*, 1162-1167.

Bernstein, N. (1967) *The Coordination and Regulation of Movement*. (London: Pergamon Press).

Bever, T. G., J. A. Fodor and M. A. Garrett. (1968) Formal limitation of associationism. In *Verbal Behavior and General Behavior Theory*, ed. by T. R. Dixon and D. L. Horton. (Englewood Cliffs, N.J.: Prentice Hall).

Bohm, D. (1957) *Causality and Chance in Modern Physics*. (New York: Harper Torchbooks).

Doty, R. W. (1968) Neural organization of deglutition. In *Handbook of Physiology*, vol. 6, section 6, ed. by C. F. Code. (Washington: American Physiological Society).

Easton, T. A. (1972) On the normal use of reflexes. *American Scientist 60*, 591-599.

Fowler, C. A. (1977) Timing control in speech production. (Unpublished Doctoral dissertation, University of Connecticut).

Gel'fand, I. M., V. S. Gurfinkel, H. L. Tsetlin and M. L. Shik. (1971) Some problems in the analysis of movements. In *Models of the Structural-Functional Organization of Certain Biological Systems*, ed. by I. M. Gel'fand, V. S. Gurfinkel, S. V. Fomin and M. T. Tsetlin. (Cambridge, Mass.: MIT Press), pp. 329-345.

Gel'fand, I. M. and M. L. Tsetlin. (1962) Some methods of control for complex systems. *Russian Mathematical Surveys 17*, 95-116.

Gel'fand, I. M. and M. L. Tsetlin. (1971) Mathematical modeling of mechanisms of the central nervous system. In *Models of the Structural-Functional Organization of Certain Biological Systems*, ed. by I. M. Gel'fand, V. S. Gurfinkel, S. V. Fomin and M. T. Tsetlin. (Cambridge, Mass.: MIT Press).

Gibson, J. J. (1950) *The Perception of the Visual World*. (Boston: Houghton Mifflin).

Gibson, J. J. (1958) Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology 49*, 182-194.

Gibson, J. J. (1961) Ecological optics. *Vision Research 1*, 253-262.

Gibson, J. J. (1966) *The Senses Considered as Perceptual Systems*. (Boston: Houghton-Mifflin).

Gibson, J. J. (1968) What gives rise to the perception of motion? *Psychological Review 75*, 335-346.

Gibson, J. J. (1977) The theory of affordances. In *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, ed. by R. E. Shaw and J. D. Bransford. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Greene, P. H. (1969) Seeking mathematical models for skilled actions. In *Biomechanics*, ed. by D. Bootzin and H. C. Muffey. (New York: Plenum Press).

Greene, P. H. (1972) Problems of organization of motor systems. In *Progress in Theoretical Biology*, vol. 2, ed. by R. Rosen and F. Snell. (New York: Academic Press).

Greenwald, A. (1970) Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review 77*, 73-99.

Grillner, S. (1975) Locomotion in vertebrates. *Psychological Review 55*, 247-304.

Hubbard, A. W. (1960) Homokinetics: Muscular function in human movement. In *Science and Medicine of Exercise and Sport*, ed. by W. R. Johnson. (New York: Harper).

Kots, Y. M. and A. V. Syrovegin. (1966) Fixed set of variants of interactions in the muscles of two joints in the execution of simple voluntary movements. *Biophysics 11*, 1212-1219.

Krinskiy, V. I. and M. L. Shik. (1964) A simple motor task. *Biophysics 9*, 661-666.

Lee, D. N. (1974) Visual information during locomotion. In *Perception: Essays in Honor of J. J. Gibson*, ed. by R. B. MacLeod and H. L. Pick, Jr. (Ithaca: Cornell University Press).

Lee, D. (1976) A theory of visual control of braking based on information about time-to-collision. *Perception 5*, 437-459.

Lee, D. and E. Aaronson. (1974) Visual proprioceptive standing in human infants. *Perception & Psychophysics 15*, 529-532.

Lishman, J. R. and D. N. Lee. (1973) The autonomy of visual kinaesthesis. *Perception 2*, 287-294.

Mace, W. (1977) James J. Gibson's strategy for perceiving: Ask not what's inside your head, but what your head's inside of. In *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, ed. by R. E. Shaw and J. D. Bransford. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Medawar, P. (1973) A geometrical model of reduction and emergence. In *Studies in the Philosophy of Biology*, ed. by F. Ayala and T. Dobzhansky. (Los Angeles: University of California Press).

Pattee, H. H. (1970) The problem of biological hierarchy. In *Toward a Theoretical Biology*, vol. 3, ed. by C. H. Waddington. (Chicago: Aldine).

Pattee, H. H. (1973) The physical basis and origin of hierarchical control. In *Hierarchy Theory: The Challenge of Complex System*, ed. by H. H. Pattee. (New York: Braziller).

Pittenger, J. and R. Shaw. (1975) Aging faces as viscal-elastic events: implications for a theory of non-rigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance 1*, 374-382.

Powers, W. T. (1973) *Behavior: The Control of Perception*. (Chicago:

Aldine).

Putnam, H. (1973) Reductionism and the nature of psychology. *Cognition 2*, 131-146.

Schiff, W. (1965) Perception of impending collision: A study of visually directed avoidance behavior. *Psychological Monographs 79* (Whole No. 604).

Schmidt, R. A. (1975) A schema theory of discrete motor skill learning. *Psychological Review 82*, 225-260.

Sessle, B. J. and A. G. Hannam. (1975) *Mastication and Swallowing: Biological and Clinical Correlates*. (Toronto: University of Toronto Press).

Shaw, R. and M. McIntyre. (1974) Algoristic foundations to cognitive psychology. In *Cognition and the Symbolic Processes*, ed. by W. Weimer and D. S. Palermo. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Shaw, R., M. McIntyre and W. Mace. (1974) The role of symmetry in event perception. In *Perception: Essays in Honor of J. J. Gibson*, ed. by R. B. MacLeod and H. L. Pick, Jr. (Ithaca: Cornell University Press).

Shik, M. L. and G. N. Orlovskii. (1976) Neurophysiology of locomotor automatism. *Psychological Review 56*, 465-501.

Sperry, R. (1952) Neurology and the mind-brain problem. *American Scientist 40*, 291-312.

Suppes, P. (1969) Stimulus-response theory of finite automatia. *Journal of Mathematical Psychology 6*, 327-355.

Trowbridge, M. H. and H. Cason. (1932) An experimental study of Thorndike's theory of learning. *Journal of General Psychology 7*, 245-258.

Turvey, M. T. (1975) Perspectives in vision: Conception or perception? In *Reading, Perception and Language*, ed. by D. Duane and M. Rawson. (Baltimore: York).

Turvey, M. T. (1977a) Contrasting orientations to a theory of visual information processing. *Psychological Review 84*, 67-88.

Turvey, M. T. (1977b) Preliminaries to a theory of action with reference to vision. In *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, ed. by R. E. Shaw and J. D. Bransford. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Turvey, M. T. and S. Prindle. (1978) Modes of perceiving: Abstracts, comments, and notes. In *Modes of Perceiving and Processing of Information*, ed. by H. Pick and E. Saltzman. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Turvey, M. T., R. Shaw and W. Mace. (in press) Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. In *Attention and Performance*, vol. 7, ed. by J. Requin. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Warren, R. (1976) The perception of egomotion. *Journal of Experimental Psychology: Human Perception and Performance 2*, 448-456.

Weiss, P. (1941) Self-differentiation of the basic pattern of coordination. *Comparative Psychology Monograph 17*, 21-96.

Wickstrom, R. L. (1970) *Fundamental Motor Patterns*. (Philadelphia: Lea & Febiger).

II. PUBLICATIONS AND REPORTS
III. APPENDIX

PUBLICATIONS AND REPORTS

Cooper, Franklin S., Paul Mermelstein and Patrick W. Nye. (1977) Speech synthesis as a tool for the study of speech production. In *Dynamic Aspects of Speech Production*, ed. by Masayuki Sawashima and Franklin S. Cooper. (Tokyo: University of Tokyo Press), 307-322.

Dorman, Michael F., Diane Kewley-Port, Susan Brady and Michael T. Turvey. (1977) Vowel recognition: Inferences from studies of forward and backward masking. *Quarterly Journal of Experimental Psychology 29*, 483-497.

Fischer, F. William, Isabelle Y. Liberman and Donald Shankweiler. (in press) Reading reversals in developmental dyslexia. *Cortex*.

Gay, Thomas. (1977) Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America 62*, 183-193.

Gay, Thomas. (1977) Cinefluorographic and electromyographic studies of articulatory organization. In *Dynamic Aspects of Speech Production*, ed. by Masayuki Sawashima and Franklin S. Cooper. (Tokyo: University of Tokyo Press), 85-105.

Harris, Katherine S. (1977) The study of articulatory organization: Some negative progress. In *Dynamic Aspects of Speech Production*, ed. by Masayuki Sawashima and Franklin S. Cooper. (Tokyo: University of Tokyo Press), 71-82.

Katz, Leonard. (1977) Reading ability and single-letter orthographic redundancy. *Journal of Educational Psychology 69*, 653-659.

Mark, Leonard S., Donald Shankweiler, Isabelle Y. Liberman and Carol A. Fowler. (1977) Phonetic recoding and reading difficulty in beginning readers. *Memory & Cognition*, vol. 5(6), 623-629.

APPENDIX

DDC (Defense Documentation Center) and ERIC (Educational Resources Information Center) numbers SR-21/22 to SR-50:

Status Report		DDC	ERIC
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	ED-077-285
SR-33	January - March 1973	AD 762373	ED-081-263
SR-34	April - June 1973	AD 766178	ED-081-295
SR-35/36	July - December 1973	AD 774799	ED-094-444
SR-37/38	January - June 1974	AD 783548	ED-094-445
SR-39/40	July - December 1974	AD A007342	ED-102-633
SR-41	January - March 1975	AD A103325	ED-109-722
SR-42/43	April - September 1975	AD A018369	ED-117-770
SR-44	October - December 1975	AD A023059	ED-119-273
SR-45/46	January - June 1976	AD A026196	ED-123-678
SR-47	July - September 1976	AD A031789	ED-128-870
SR-48	October - December 1976	AD A036735	ED-135-028
SR-49	January - March 1977	AD A041460	ED-141-864
SR-50	April - June 1977	AD A044820	ED-144-138
SR-51/52	July - December 1977	AD A049215	ED-147-892
SR-53, v. 1	January - March 1978	**	**
SR-53, v. 2	January - March 1978	**	**

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

**DDC and/or ERIC order numbers not yet assigned.

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Haskins Laboratories, Inc.
270 Crown Street
New Haven, Connecticut 06510

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

N/A

3. REPORT TITLE

Haskins Laboratories Status Report on Speech Research, No. 53, vol. 1, January - March, 1978.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)

Staff of Haskins Laboratories; Alvin M. Liberman, P.I.

6. REPORT DATE

March 1978

7a. TOTAL NO. OF PAGES

172

7b. NO. OF REFS

197

8a. CONTRACT OR GRANT NO.

HD-01994
V101(134)P-342
N01-HD-1-2420
RR-5596
BNS76-82023
MCS76-81034
NS13870
NS13617

9a. ORIGINATOR'S REPORT NUMBER(S)

SR-53, vol. 1 (1978)

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES

N/A

12. SPONSORING MILITARY ACTIVITY

See No. 8

13. ABSTRACT

This report (1 January - 30 April) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

- -Articulatory Synthesis - A Tool for the Perceptual Evaluation of Articulatory Gestures
- -On the Relation between Processing the Roman and the Cyrillic Alphabets: A Preliminary Analysis with Bi-alphabetical Readers
- -Bi-alphabetical Lexical Decision
- -Lexical Decision for Inflected Nouns
- -Phonetic Plausibility of the Segmentation of Tones in Thai Phonology
- -Closure Hiatus: Cue to Voicing, Manner and Place of Consonant Occlusion
- -Metaphoric Comprehension: Studies in Reminding and Resembling
- -Skill Acquisition: An Event Approach with Special Reference to Searching for the Optimum of a Function of Several Variables

DD FORM 1 NOV 65 1473 (PAGE 1)

UNCLASSIFIED
Security Classification

S/N 0101-807-6811

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

A-31408

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Articulatory Synthesis; Speech evaluation						
Alphabet; Roman, Cyrillic, Reading						
Lexical decisions, Bi-alphabetic						
Lexical decisions, nouns						
Tones, Segmentation in Thai						
Closure; Voicing Manner and Place Cues						
Metaphor; Comprehension						
Skill Acquisition; Optimization						

DD FORM 1 NOV 65 1473 (BACK)

S/N 0101-807-6821

UNCLASSIFIED

Security Classification

A-31409

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 January - 31 March 1978

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994

Assistant Chief Medical Director for Research and Development,
Research Center for Prosthetics, Veterans Administration
Contract V101(134)P-342

National Institutes of Child Health and Human Development
Contract N01-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-5596

National Science Foundation
Grant BNS76-82023

National Science Foundation
Grant MCS76-81034

National Institute of Neurological and Communicative
Disorders and Stroke
Grant NS13870
Grant NS13617

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper, Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey, Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Thomas Baer
Fredericka Bell-Berti
Gloria J. Borden*
Guy Carden*
Robert Crowder*
Steven B. Davis
Michael Dorman*
Donna Erickson
William Ewan*
Carol A. Fowler*
Frances J. Freeman*
Jane H. Gaitenby
Thomas J. Gay*
Katherine S. Harris*
Alice Healy*
David Isenberg
Leonard Katz*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann
Ignatius G. Mattingly*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Donald P. Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Robert Verbrugge*
Hirohide Yoshioka[1]

Technical and Support Staff

Eric L. Andreasson
Elizabeth P. Clark
Donald Hailey
Terry Halwes
Elly Knight*
Sabina D. Koroluk
Agnes McKeon
Nancy R. O'Brien
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
Bobbe Wire
David Zeichner

Students*

Linda D'Antonio
William Balch
Steve Braddon
David Dechovitz
Laurel Dent
Susan Lea Donald
F. William Fischer
Hollis Fitch
Carole E. Gelfer
Janette Henderson
Nieba Jones
Morey J. Kitzman
Andrea G. Levitt
Roland Mandler
Leonard Mark
Nancy McGarr
Georgia Nigro
Mary Jo Osberger
Abigail Reilly
Arnold Shapiro
Emily Tobey
Betty Tuller
Harold Tzeutschler
Michele Werfelman

---

*Part-time

[1] Visiting from University of Tokyo, Japan.

CONTENTS

I. Manuscripts and Extended Reports

Detection Errors on *The* and *And*: Evidence for Reading Units Larger than the Word -- Adam Drewnowski and Alice F. Healy 1

Some Observations on the Perception of [s]+Stop Clusters -- Peter J. Bailey and Quentin Summerfield 25

Perceptual Integration of Cues for Stop, Fricative and Affricate Manner -- Bruno H. Repp, Alvin M. Liberman, Thomas Eccardt and David Pesetsky 61

Diachronic Tone Splits and Voicing Shifts in Thai: Some Perceptual Data -- Arthur S. Abramson and Donna M. Erickson 85

A Range Effect in the Perception of Voicing -- Susan A. Brady and Christopher J. Darwin 97

A Note on Perceptuo-Motor Adaptation of Speech -- Quentin Summerfield, Peter J. Bailey and Donna Erickson 105

Interdependence of Voicing and Place Decisions for Stop Consonants in Initial Position -- Bruno H. Repp 117

II. Publications and Reports

III. Appendix: DDC and ERIC numbers (SR-21/22 - SR-50)

I. MANUSCRIPTS AND EXTENDED REPORTS

Detection Errors on The and And: Evidence for Reading Units Larger than the Word[*]

Adam Drewnowski[†] and Alice F. Healy[††]

ABSTRACT

In five experiments subjects read 100-word passages and circled instances of a given target letter, letter group, or word. In each case subjects made a disproportionate number of detection errors on the common function words the and and. The predominance of errors on these two words was reduced for passages in which the words were placed in an inappropriate syntactic context and for passages in which word-group identification was disturbed by the use of mixed type-cases or a list, rather than a paragraph, format. These effects for the word and were not found for the control word ant. These results were taken as evidence that familiar word sequences may be read in units larger than the word, probably short syntactic phrases or word frames. A tentative model of the reading process consistent with these results is proposed.

INTRODUCTION

The present study employs a detection task to investigate the possibility that high frequency words may be read in terms of units larger than the word--word frames, phrases or syntactic groups. It has been observed that subjects searching for instances of a given target letter in printed text make a disproportionate number of errors on the word the (Corcoran, 1966; Healy, 1976). Healy found the high frequency of the word the to be critical, in support of the view that frequent words are read in terms of units larger than

---

[*] The paper appears in Memory & Cognition, vol. 5.

[†] The Rockefeller University, New York.

[††] Also Yale University, New Haven.

Acknowledgment: Order of authorship was determined alphabetically. This research was supported in part by U.S. Public Health Service Grant MH 23878 to The Rockefeller University, PHS Grants MH 26573 and RR 07015 to Yale University, and NICHD Grant HD 01994 and BRSG Grant RR 05596 to Haskins Laboratories. We are indebted to the Deans of Cornell and Mount Sinai Medical Schools for their assistance in providing subjects for Experiments I, II, and IV, to W. Brown, J. Gottlieb and W. S. Terry for help with the data analyses of Experiment III, and to W. K. Estes, C. Lee and L. Marks for helpful discussions about this research.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 2 (1978)]

the letter, for example, spelling patterns (Gibson, 1965) or vocalic center groups (Spoehr and Smith, 1973). Here we extend that experimental paradigm by using words rather than letters as targets.

We propose to test a specific set of hypotheses that may enter into a model of the reading process. In common with other investigators, we shall assume that reading involves the use of a hierarchy of graphological, orthographic, lexical, and syntactic processing skills (Gibson, 1971; LaBerge and Samuels, 1974; Doehring, 1976; Estes, in press). We distinguish, in particular, between five separate levels of processing written text, which we define in terms of the units available at each level--letters, letter groups, words, phrases and larger units such as clauses or sentences. We assume further that the completion of processing at a given level is tantamount to the identification of the unit at that level. Identification can be monitored through a detection task requiring subjects to circle every instance of a given target, which can be at any one of the processing levels. In previous studies (for example, Corcoran, 1966; Healy, 1976), only letters were used as targets, but in the present study targets can be letters, letter groups, or words.

In the formulation of our model we make the further assumption that, in the course of normal reading, subjects tend to process stimuli at the highest level available to them, in parallel with processing at lower levels. This assumption of parallel processing contrasts to that made by Gibson (1971) who postulates that subjects move through the linguistic hierarchy sequentially. We prefer the parallel assumption, or at least the weaker assumption (Estes, in press) that processing at higher levels can begin before processing at some lower level is complete, because we know from the study of Healy (1976) that subjects fail to detect letters when word units become available. This result also implies that once a unit has been identified at any given level, subjects will proceed to the next unit of text at that level without necessarily completing the processing at the lower levels. For example, once a word has been identified, the subject may move on to the next word without necessarily identifying all the letters and letter groups within the word. Consequently, when subjects are searching for targets at a given level, we should expect them to make more detection errors when they are able to identify units at a level higher than the level of the target than when they are able to identify only units at the target level or below.

The highest level of processing reached will no doubt partly depend on the subject's reading skill (Gibson, 1971; LaBerge and Samuels, 1974), but it will also depend on the nature of the stimulus materials used and the task demands (Gibson, 1971; Estes, 1975). In the present study, we therefore employ prose passages, where the highest level of processing is at least as large as the phrase, and we compare these to scrambled-word passages where the highest level of processing is the level of the word. Furthermore, we consider passages, such as the scrambled-letter passage of Healy (1976), where the highest level of processing is the level of the letter. In other words, in the detection tasks of the present study, we independently manipulate the level of the target and the processing levels available in the search passage.

If processing at a higher level is in some way impeded, processing at the lower levels will be more likely to proceed to completion, resulting in better detection of targets at the lower levels. Conversely, if processing at a higher level is in some way facilitated, processing at the lower levels should be less likely to proceed to completion. Familiarity with a unit at a given level will presumably facilitate processing of the unit at that level. Thus the use of words of high frequency should facilitate processing, or identification, at the word level, and the use of commonly encountered syntactic phrases or word frames should facilitate processing at the phrase level. Thus, for example, subjects may be able to identify a familiar phrase before identification of the specific words in the phrase is completed. On the other hand, when phrases are not familiar, the subjects should not be able to complete processing at the phrase level before completing processing at the word level.

Consequently, we introduce variations in the search passage that impede or promote the formation of higher level units. In addition, we examine the detection of targets in frequent and in infrequent words, and the detection of targets in syntactically correct phrases and in comparable syntactically incorrect word groups. We also introduce a modification of the detection task that requires the subjects to attend to the meaning of words, as in normal reading.

EXPERIMENT I

The results of the study by Healy (1976) suggest, in accord with our proposed model, that subjects fail to complete processing at the letter level when processing at the word level is facilitated. This conclusion is based on the finding of a disproportionate number of detection errors on a given letter (*t*) when that letter was embedded in a frequent word (*the*). The present study aims to determine whether, in analogy with this finding and in accord with our model, subjects fail to complete the processing at the word level when processing at the phrase level is facilitated. Specifically, we are led to predict that a disproportionate number of detection errors will occur on a given word when that word is embedded is a familiar phrase. In order to test this possibility we use as targets both the letter *t* and the letter group *the*, and we employ search passages so constructed that the letter *t* is always part of the letter group *the*. This technique enables us to make a direct comparison of performances on the letter and letter-group detection tasks. Since we expect subjects to make a disproportionate number of detection errors on the letter group *the* when it occurs as the word *the* contained in a familiar phrase, we examine two passages--a prose passage, where every instance of the word *the* necessarily occurs in a syntactically appropriate phrase, and a scrambled-word passage so constructed that only half of the occurrences of the word *the* occur within syntactically appropriate word phrases.

Method

Subjects. Sixty-four male and female students at Cornell Medical School, who were attending a neuroanatomy lecture, served as subjects in a group experiment conducted in the classroom. They were divided into two groups. Thirty-four of the subjects were in Group T and thirty in Group The.

Design and Materials. Two 100-word passages, typed on separate sheets of paper, were constructed for the present experiment. One passage, herafter referred to as the "prose passage," contained 12 instances of the word *the*, 24 words that included the letter string *the* but no other instance of the letter *t* (examples: ba*the*d and ra*the*r), and 64 filler words chosen with the restriction that no word included the letter *t*. Every instance of the letter *t* in the passage was thus part of the letter string *the*. (See Appendix for a copy of the prose passage.)

The second passage, hereafter referred to as the "scrambled-word" passage, was derived from the prose passage. The 12 *thes* and the 24 words containing the letter string *the* were in the same locations as in the prose passage, and the punctuation marks remained the same. The order of the 64 filler words was random, with the single constraint that out of the 12 instances of the word *the*, six were followed by nouns (appropriate context), and six by other parts of speech (inappropriate context).

Procedure. The subjects received written instructions that differed for the two groups. Subjects in Group T were asked to circle instances of the letter *t* as target, while subjects in Group The were asked to circle instances of the letter group *the*, either by itself, or embedded in another word. Both groups were told to read each passage at their normal reading speed and to encircle each instance of the target with a pen or a pencil. The subjects were told that if they ever realized that they had missed a target, they should not retrace their steps to encircle it. They were told further that they were not expected to detect all targets, so they should not slow their normal reading speed in order to be overcautious about encircling the targets. Each subject was shown both passages. Half the subjects in each group were shown the prose passage first; the other half were shown the scrambled-word passage first. Subjects were told to read the two passages in the order in which they were stapled together and to go on to the second passage as soon as they had finished the first.

Results

The results of the present experiment are summarized in Table 1, which includes for the two passages the means and the standard errors of the means for the total number of errors, for the number of errors on the word *the*, and for the conditional percentage of detection errors on the word *the* given an error. Means and standard errors were derived by computing scores for each subject and then averaging across subjects. All errors considered were omission errors (misses), since there were virtually no false alarm errors in any of the present experiments. Consequently, the mean total error score is the sum of the mean error score on the word *the* and the mean error score on the words containing the embedded letter string *the*. The conditional percentage of detection errors on the word *the* was derived for a given subject by determining the ratio of the number of errors on the word *the*, divided by the total number of errors. By chance alone, the conditional percentage of errors on the word *the* should be 33.3, since 12 of the 36 targets involve the word *the*. Healy (1976) found this conditional percentage to be the most sensitive index of performance in this situation, since it is not influenced by the speed-accuracy tradeoff typically found in such a task. Further, analyses of the present data revealed no significant difference in conditional percentages between subjects with high total error scores (10 or above) and subjects with low error scores. Because the conditional error percentages constitute the

TABLE 1: Means and standard errors of means (in parentheses) for error frequencies and conditional percentages for Groups T and The of Experiment I.

				Errors on word *the*			Errors on word *the* given error	
Group	N	Passage	Total errors	Total	appropriate context	inappropriate context	%	N'
T	34	Prose	7.44 (.90)	5.67 (.73)	---	---	72 (4)	32
		Scrambled word	5.65 (.86)	3.88 (.67)	2.38 (.36)	1.50 (.34)	65 (5)	29
The	30	Prose	5.63 (.69)	3.56 (.52)	---	---	60 (6)	30
		Scrambled word	4.50 (.77)	2.66 (.47)	1.53 (.30)	1.13 (.21)	61 (5)	24

Note. In this table and in the succeeding tables in this paper, the total number of subjects (N) does not necessarily equal the number of subjects on which the mean conditional percentages are based (N'), since not all subjects made errors on each passage.

critical dependent variable, they are emphasized throughout our discussion of the results, although total error scores for all experiments are also shown in the tables below along with error scores for the word *the*.

"T" and "The" Detection Tasks. The conditional error percentages were not significantly lower for Group The than for Group T on either the prose [$t(60) = 1.54$, $p > .10$] or the scrambled-word passages [$t(51) = .46$, $p > .10$]. Both values for Group The were in fact significantly above chance level (33 percent): prose: $t(29) = 4.58$, $p < .001$; scrambled word: $t(23) = 5.83$, $p < .001$. This similarity between Group T and Group The indicates that even those subjects who were specifically instructed to search for instances of the letter group *the* made a disproportionate number of detection errors whenever *the* occurred on its own, rather than as part of another word. Although these data, obtained with both prose and scrambled-word passages, appear counterintuitive, they are consistent with our proposed set of hypotheses. These data are equally consistent with Corcoran's (1966) hypothesis that the word *the*, being redundant (that is, predictable from the prior word context), may not be scanned by the reader. However, the redundancy hypothesis would predict considerably fewer errors on the word *the* in the scrambled-word passage, where its occurrence cannot be predicted on the basis of prior context. This result was not obtained; hence it seems unlikely that the word *the* is not scanned by the reader. We propose instead that the word *the* is scanned, but that the processing at the phrase level is completed before the word *the* itself can be fully processed and identified. Phrase-level units, which are clearly available in the prose passage, might also be formed in the scrambled-word passage whenever the target word *the* occurs in an appropriate syntactic context. An analysis of context effects, between and within passages follows.

Prose Versus Scrambled Words. The conditional percentage of errors on the word *the* given an error was somewhat higher for the prose passage than for the scrambled-word passage for Group T, though not significantly so [$t(27) = 1.36$, $p > .10$]. For Group The, the conditional error percentages for the prose and scrambled-word passages were approximately equal. However, significantly more unconditionalized errors on the word *the* were made on the prose passage than on the scrambled-word passage by both Group T [$t(33) = 3.80$, $p < .001$] and Group The [$t(29) = 2.44$, $p < .05$]. These data are consistent with those of Healy (1976), who reported a significant difference in unconditionalized errors, but not in conditional percentages between comparable prose and scrambled-word passages.

It is important to consider the possibility that the prose passage employed is semantically odd in ways that might lead subjects to spend a disproportionate amount of processing time, and hence make disproportionately fewer errors, on embedded *thes*. Many of the content words involving the embedded *thes* are peculiar, relative to the surrounding context. For example, mothers rarely discuss "psychotherapy" and "anesthesia." However, the fact that we obtained disproportionately fewer errors on the embedded *thes* for the scrambled-word passage as well as for the prose passage, suggests that the semantic oddity of the prose passage was not critical to the present results, since all words would be unexpected, or peculiar relative to the surrounding context in the scrambled-word passage. Furthermore, as we noted above, similar results were obtained by Healy (1976) for letter detection errors with another scrambled-word passage and a prose passage selected from Golding's *Lord of the Flies*.

Within-Passage Context Effects. In the scrambled-word passage, both Group T and Group The subjects made more detection errors on instances of the word *the* that were immediately followed by nouns than on those that were followed by other parts of speech. This difference between the "appropriate context" and "inappropriate context" *thes* was highly significant for Group T [$t(33) = 4.04$, $p < .001$], and was in the right direction, though not significant, for Group The [$t(29) = 1.65$, $p > .10$]. These findings of local context effects provide further support for the hypothesis that under appropriate circumstances, phrase-level units may be identified in scrambled-word passages and may impair identification, and hence detection, of lower level units.

EXPERIMENT II

In Experiment I, subjects searching for both types of targets (*t* and *the*) made more errors on the word *the* whenever it appeared in an appropriate context. However, the appropriateness of the context was confounded with such variables as the location of the target in the search passage and the nature of the words surrounding each target. The present experiment addresses the question of context effects more systematically. We now manipulate the nature of the context by altering the sequence of two filler words around each of the 12 *the* targets in two newly-constructed scrambled-word passages. Whereas in one passage, all instances of the word *the* occur in what is deemed an "appropriate" context, in the other passage, all instances of the word *the* occur in an "inappropriate" context. If processing at the phrase level impairs processing at the letter and word levels, we would expect more errors and higher conditional error percentages on *thes* in an appropriate context than in an inappropriate context.

Method

Subjects. Forty-eight male and female students at the Mt. Sinai Medical School, who were attending a biochemistry lecture, served as subjects in a group experiment conducted in the classroom. Twenty-four subjects were in Group T and twenty-four subjects were in Group The.

Design and Materials. The scrambled-word passage used in Experiment 1 was used again in the present experiment. In this passage, six instances of the word *the* were followed by nouns and six by other parts of speech. Two new passages, referred to as the "local-context" and "no-context" passages respectively, were derived from the scrambled-word passage. Both contained the same 36 targets, including 12 instances of the word *the* and 24 target words containing *the*. The same filler words and the same punctuation were used as in the scrambled-word passage. The no-context passage differed from the scrambled-word passage in one respect only: the sequence of words was partially rearranged so that no instance of the word *the* was followed by a noun. In contrast, each of the twelve *the* targets in the local-context passage was part of a meaningful syntactic group, typically a prepositional phrase. We accomplished this effect by reversing the sequence of the two filler words on either side of each target word *the*, so that a meaningless sequence of words in the no-context passage (for example, "air the of") would effectively become a syntactic group in the local-context passage (for example, "of the air"). The word *the* thus always appeared in an appropriate context in the local-context passage, and in an inappropriate context in the no-context passage.

A fourth passage, also 100 words long and constructed according to similar principles as the scrambled-word passage, was included in the present experiment. We used this passage largely to provide variety, and the results obtained will not be reported here.

Procedure. All subjects were shown all four passages, with the order of presentation being roughly counterbalanced across subjects. As in Experiment I, subjects received written instructions that differed for the two groups. Subjects in Group T were instructed to circle instances of the letter t, while subjects in Group The were to circle instances of the letter group the. The procedure was otherwise identical to that used in Experiment I.

Results

The results are summarized in Table 2, which includes for the three passages the means and the standard errors of the means for the total number of errors, for the number of errors on the word the, and for the conditional percentage of errors on the word the given an error.

TABLE 2: Means and standard errors of means (in parentheses) for error frequencies and conditional percentages for Groups T and The of Experiment II.

Group	N	Passage	Total errors	Errors on word the	Errors on word the given error %	N'
T	24	scrambled word	5.95 (.85)	4.16 (.65)	69 (5)	24
		no context	5.00 (1.05)	2.79 (.81)	46 (8)	20
		local context	6.83 (.94)	5.00 (.78)	69 (5)	22
The	24	scrambled word	5.12 (.78)	3.12 (.60)	61 (7)	19
		no context	2.95 (.62)	1.04 (.51)	18 (6)	21
		local context	5.16 (.86)	3.21 (.52)	51 (8)	22

The conditional percentage of errors on the word *the* given an error was substantially higher for the local-context passage than for the no-context passage for subjects in both Group T [$t(19) = 2.60$, $p < .05$] and Group The [$t(20) = 5.18$, $p < .001$]. For the local-context passage, the conditional error percentage was significantly above chance level for both groups [Group T: $t(21) = 6.59$, $p < .001$; Group The: $t(21) = 2.40$, $p < .05$], whereas for the no-context passage it was not different from chance level for Group T [$t(23) = 1.54$, $p > .10$] and was significantly below chance level for Group The [$t(20) = 2.51$, $p < .05$]. Detection errors on the word *the* are thus substantially decreased whenever the word *the* is removed from its habitual context. This effect is observed regardless of whether the subjects are searching for the letter *t*, or for the word *the*. As both passages used are scrambled-word passages, the context effects observed do not depend on the presence of larger syntactic or semantic units such as clauses or sentences.

As for the scrambled-word passage, the conditional percentage of errors on the word *the* given an error was significantly above chance level for both Group T [$t(23) = 7.13$, $p < .001$] and Group The [$t(18) = 4.17$, $p < .001$]. More detection errors on the word *the* were made by subjects in both Group T [$t(23) = 3.49$, $p < .01$] and Group The [$t(23) = 7.03$, $p < .001$] when the word *the* preceded a noun than when it did not. The mean number of errors on the word *the* was 2.70 for *thes* in an appropriate context and 1.45 for *thes* in an inappropriate context for subjects in Group T, and 2.20 and .91 respectively for subjects in Group The. These results for the scrambled-word passage essentially replicate those obtained for the same passage and different subject population in Experiment I.

EXPERIMENT III

Experimen[illegible] and II demonstrate the effects of word context on the detection of [illegible] target *t* as well as on the detection of the target *the*. Subjects make [illegible]e errors on the word *the* when it appears in an appropriate context than when it appears out of context, and we have suggested that the word *the* may be processed as part of a phrase-level unit. We now attempt to impede the formation of phrase-level units by using purely perceptual variables, instead of syntactic or semantic variables, such as were used in Experiment II. For this purpose we employ two types of manipulations--one involves the use of mixed type-cases (see Fisher, 1975 and McClelland, 1976, for other examples of the use of this variable), and the second involves a change in the passage layout. Two mixed type-case passages were constructed. In the first passage (mixed letter), every other letter is typed in capitals in order to disturb word identification and make it more difficult for the subjects to process units at levels higher than the letter. In the second passage (mixed word), every other word is typed in capitals in order to disturb phrase identification and make it more difficult for the subjects to process units at levels higher than the word. We changed the passage layout by typing the words in five vertical columns, which the subjects were instructed to read from top to bottom. This final manipulation (list passage) not only disturbs the identification of units larger than the word, but also forces the subjects to abandon their usual left-to-right reading pattern. The processing at levels higher than the word should, therefore, be virtually eliminated by the list passage. In each of these passages where the formation of units larger than the word is disturbed, our hypotheses lead us to expect a decrease in the conditional error percentages on the word *the* relative to the standard scrambled-word passage.

The present study investigates one additional question. It may be that subjects presented with scrambled-word passages do not read them in the same manner as they would a prose passage. In particular, it may be that the detection task predominates over the reading task so that subjects do not attend to the meaning of words. In order to ensure that subjects do attend to word meanings, the subjects in the present experiment are given the task of underlining every name of a living thing in each passage, in addition to their tasks of reading and circling instances of the target letter or target letter group.

Method

Subjects. One hundred and thirty-seven male and female undergraduate students of Yale University, who were taking a course in introductory psychology, served as subjects in a group experiment conducted in the classroom. The subjects were divided into two groups, with 71 subjects in Group T and 66 subjects in Group The.

Design and Materials. We employed five passages of scrambled words. Four of the five passages were identical in terms of the words used; they differed only in the format in which they were typed. The first passage (scrambled-word passage) was identical to that used in Experiments I and II. The second passage (mixed-letter passage) differed from the scrambled-word passage only in that every other letter was typed in upper case. There were two versions of the mixed-letter passage. In Version A the first letter in the passage was typed in lower case, whereas in Version B the first letter in the passage was typed in upper case. Thirty-six subjects in Group T and thirty-one subjects in Group The were shown Version A of the passage, and thirty-five subjects in Group T and thirty-five subjects in Group The were shown Version B. The third passage (mixed-word passage) differed from the first in that every other word was typed in upper case. There were two versions of the mixed-word passage. In Version A the first word in the passage was typed in lower case, and in Version B the first word in the passage was typed in upper case. Thirty-six subjects in Group T and thirty-two subjects in Group The were shown Version A, and thirty-five subjects in Group T and thirty-four subjects in Group The were shown Version B. The fourth passage (list passage) contained the same words as the other passages, but they were typed in five vertical columns, with each word typed flush left. The order of the words and the type-case of the letters were the same as in the scrambled-word passage, but all commas, periods and quotation marks were removed. The fifth passage resembled the scrambled-word passage in both construction and layout, but different sets of words were used in the two passages. This passage was employed to provide variety, so that the subjects would be less likely to realize that the four critical passages contained the same words. The results for the fifth passage will not be discussed in the present paper.

Procedure. All subjects were shown all five passages, with the fifth passage always shown in the third position. The order of the other four passages was roughly counterbalanced across subjects. Subjects received written instructions to circle instances of the letter t (Group T) or to circle instances of the letter group the (Group The). In addition, unlike previous experiments, subjects in both groups were told to underline every name of a living thing in each passage. The subjects were also told to note that one passage (list passage) would consist of five vertical columns of words and that they were to read each column of words from top to bottom. The procedure was otherwise analogous to that used in Experiments I and II.

Results

The results of the present experiment are summarized in Table 3, which includes for each of the four critical passages and each of the two groups, the means and the standard errors of the means for the number of total errors, the number of errors on the word the and the conditional percentage of errors on the word the given an error.

TABLE 3: Means and standard errors of means (in parentheses) for error frequencies and conditional percentages for Groups T and The of Experiment III.

Group	N	Passage	Total errors	Errors on word the	Errors on word the given error %	N'
T	71	scrambled word	5.18 (.45)	2.94 (.32)	55 (4)	67
		mixed letter	3.44 (.43)	1.48 (.23)	40 (4)	59
		mixed word	4.85 (.51)	2.04 (.25)	42 (4)	63
		list	2.51 (.39)	.70 (.15)	28 (5)	54
The	66	scrambled word	5.47 (.64)	2.45 (.26)	48 (4)	62
		mixed letter	3.68 (.48)	1.61 (.19)	54 (5)	54
		mixed word	4.42 (.46)	1.73 (.21)	42 (4)	61
		list	2.66 (.32)	.45 (.10)	17 (4)	51

Subjects in Group T, who were instructed to search for *ts*, showed a similar pattern of results to those in Group The, who were instructed to search for *thes*. As in Experiments I and II, the conditional error percentages for the scrambled-word passage were significantly above chance [Group T: $t(66) = 5.63$, $p < .001$; Group The: $t(61) = 3.95$, $p < .001$]. In comparison to the scrambled-word passage, the conditional percentages for the mixed-letter passage, where every other letter was capitalized, were significantly reduced for Group T [$t(56) = 3.41$, $p < .01$], though not significantly changed for Group The [$t(50) = 1.86$, $.10 < p < .05$]. The results for Group T suggest that when the processing of units at levels higher than the letter is impeded, the percentages of letter detection errors on the word *the* are reduced to near chance level. Furthermore, the conditional error percentages for the mixed-word passage, where every other word was capitalized, were significantly reduced relative to the scrambled-word passage in Group T [$t(58) = 2.56$, $p < .05$], and reduced but not significantly so in Group The [$t(58) = 0.98$, $p > .10$]. In both groups, however, the conditional percentages in the mixed-word passage remained significantly above chance [Group T: $t(62) = 2.42$, $p < .05$; Group The: $t(60) = 2.05$, $p < .05$]. The conditional error percentages for the list passage fell below chance level and were significantly lower than those in the scrambled-word passage for both groups of subjects [Group T: $t(51) = 4.35$, $p < .001$; Group The: $t(47) = 5.69$, $p < .001$]. These observations for the mixed-word and list passages indicate that even when the processing of units at levels higher than the word is impeded, the percentage of detection errors on the word *the* is reduced.

An analysis of the underlining task revealed both a high level of performance and no significant differences in performance levels between the two groups of subjects or among the four passages. Excluding the word *I*, there were six instances of names of living things in each passage. The mean percentage of misses on these targets was 19.1, and the mean percentage of false alarms was 1.0. Consequently, we are satisfied that subjects did consider the meaning of the words during the detection task. The results for the scrambled-word passage suggest, therefore, that a large preponderance of detection errors on the word *the* is found even when the subjects consider the meaning of the words. In addition, the results for the scrambled-word passage compared to those for the other three passages suggest that the majority of detection errors on the word *the* is due to the processing at levels higher than words. Impeding the processing at these higher levels results in the observed decrease in conditional percentages of errors on the word *the*. The present experiment used perceptual variables to achieve this result. Note that Corcoran's (1966) redundancy hypothesis, which postulates that the word *the*, being redundant, is not scanned, cannot give a simple account of the present pattern of results. The predictability of the word *the* does not change with a change in type-case or passage layout; nevertheless changes were found in the percentages of detection errors on the word *the* with such changes in passage format.

EXPERIMENT IV

We aim to extend the generality of the results obtained in the preceding experiments and in the study by Healy (1976) to another target letter and another high-frequency word. The word *and*, the third most frequent word in the English language (Kučera and Francis, 1967) and the target letter *n* were accordingly selected for study. We constructed a new scrambled-word passage in which every instance of the letter *n* occurred within the letter group *and* and an equivalent control passage in which all occurrences of the word *and* and of words containing the embedded letter string *and* were replaced by the word *ant* and by words containing the embedded letter string *ant*, respectively. The control passage provides a powerful test both of the frequency hypothesis (Healy, 1976) and of the hypothesis that the subject may decide not to scan fully a given word on the basis of its global features as detected in peripheral vision (for example, Hochberg, 1970). Both trigrams (*and* and *ant*) are equal in length, share the initial two letters, and have a similar word-shape. Additional advantages of this comparison are the virtually identical embedded trigram frequencies of the two letter strings (Underwood and Schulz, 1960) and the fact that the letter *n* occurs in the same location and is pronounced similarly in both cases. The word *ant* has the further advantage of not being archaic, unlike the word *thy* employed by Healy (1976) in an equivalent control condition. The two words (*and* and *ant*) differ only in frequency, with *and* being by far the more frequent, and in part of speech, with *ant* being a concrete noun rather than a function word.

In addition, the importance of the location of the target letter (Corcoran, 1966) was tested by the use of a control passage of scrambled letters analogous to that employed by Healy (1976). Finally, the processing of text at levels higher than the word was impaired by the use of a list passage analogous to that employed in Experiment III.

Following the earlier results, we expect conditional error percentages to be above chance on the *and* passage and at chance level on the *and* list, the *ant* passage and the scrambled-letter passage.

Method

Subjects. Twenty-four male and female students at Mt. Sinai Medical School, who were attending a biochemistry lecture, served as subjects in a group experiment conducted in the classroom.

Design and Materials. We constructed four new 100-word passages. One passage, hereafter referred to as the "scrambled-word *and* passage" included 12 instances of the word *and*, 24 words that included the letter string *and* but no other instance of the letter *n*, and 64 filler words chosen from an article in *The New York Times*, with the restriction that no word included the letter *n*. These constraints ensured that every instance of the letter *n* in the passage was part of the letter string *and*. The order of the words within the scrambled-word passage was random, with the single constraint that six out of the twelve instances of the word *and* occurred between like parts of speech (appropriate context), whereas the remaining six occurred between unlike parts of speech (inappropriate context). Punctuation marks were inserted arbitrarily.

The second passage, hereafter referred to as the "scrambled-letter passage," was derived from the scrambled-word passage described above. In order to form the scrambled-letter passage, the scrambled-word passage was divided into 20 consecutive five-word groups, and the letters within each of the 20 groups were randomized. A given letter thus did not necessarily remain within the same word, but did remain within the same word group. The *n*'s, punctuation marks, and "interword" spaces were kept in the same locations as in the scrambled-word passage.

The third passage, referred to as the "scrambled-word *ant* passage," was also derived from the first. It was identical to the scrambled-word *and* passage in every respect, except that every instance of the word *and* was replaced by the word *ant*, and every target word containing the letter string *and* was replaced by another target word containing the letter string *ant*. The two classes of words containing the target letter string (those containing *and* and those containing *ant*) were roughly matched for number of letters, number of syllables, number of vocalic center groups, frequencies according to Kučera and Francis, and position of the letter string within the word. (For example, the word h*and*le was replaced by the word p*ant*ry.) Since the locations of *ant* targets precisely matched those of *and* targets, the "appropriateness" of the context of the *ant* targets did not match that of *and* targets. In fact, all *ant* targets probably were in "inappropriate" contexts. The two scrambled-word passages therefore differ not only in the frequency and the nature of the target, but also in the "appropriateness" of the context surrounding that target.

The fourth passage, hereafter referred to as the "list passage," was also derived from the first. The 100 words were typed in five vertical columns of 20 words each. The order of the words and the type-case of the letters were the same as in the scrambled-word *and* passage, but all commas and periods were removed.

Procedure. The procedure was similar to that used in Experiments I and II. Subjects received written instructions to read each passage and encircle each instance of the letter *n*. Each subject was shown all four passages, with the order of the four passages counterbalanced across subjects.

Results

The results of the present experiment are summarized in Table 4, which includes for the four passages the means and the standard errors of the means for the number of total errors, for the number of errors in "*and*/*ant* locations," and for the conditional percentage of errors in "*and*/*ant* locations" given an error. An error in an *and*/*ant* location is defined as an error in the word *and* in both the scrambled-word *and* passage and the list passage, the word *ant* in the scrambled-word *ant* passage, or in the corresponding locations in the scrambled-letter passage. (Recall that the *n*'s were in the same locations in the scrambled-word and scrambled-letter passages.) As in Experiments I-III, the chance conditional percentage of errors in *and*/*ant* locations is 33 percent, since 12 of the 36 *n*'s are in the words *and* or *ant* in each scrambled-word passage and in the list passage.

TABLE 4: Means and standard errors of means (in parentheses) for error frequencies and conditional percentages of Experiment IV.

Passage	Total errors	Errors in *and*/*ant* locations	Errors at *and*/*ant* locations given error %	N'
scrambled word *and*	7.12 (1.06)	5.75 (.79)	84 (2)	20
scrambled letter	5.20 (1.03)	2.54 (.55)	55 (5)	21
scrambled word *ant*	3.25 (.62)	.71 (.22)	24 (7)	19
list	1.96 (.50)	.91 (.34)	34 (9)	13

Note: Total N = 24.

Scrambled Word Versus Scrambled Letter. The mean conditional percentage of errors on the word *and* in the scrambled-word *and* passage was significantly above chance level [$t(19) = 19.69$, $p < .001$] and significantly above the conditional error percentage for the scrambled-letter passage [$t(19) = 5.89$, $p < .001$].

These data, which confirm the results obtained by Healy (1976, Experiment 1), are inconsistent with the hypothesis that the location of the target letter may account for the preponderance of errors on the word *and*, since the *n*'s are in the same locations with respect to the word boundaries in the two passages. However, whereas Healy found the conditional percentage of errors at *the* locations to be significantly less than chance in the scrambled-letter passage, the conditional percentage of errors in *and* locations in the present scrambled-letter passage is significantly greater than chance [$t(20) = 3.78$, $p < .01$]. A possible explanation for this discrepancy is suggested by Corcoran's (1966) finding that the position of the letter in the word does have some effect on detection probability, and that later letters (in his case *e*'s) are more likely to be missed than early ones.

And Versus Ant Passage. Whereas the conditional percentage of errors at *and*/*ant* locations given an error was significantly above chance for the *and* passage, it was not significantly different from chance [$t(18) = 1.25$, $p > .10$] for the *ant* passage. This observation, which is inconsistent with the location and pronunciation hypotheses since the location and pronunciation of the letter *n* are the same in both passages, provides support for the hypothesis that the frequency of a unit at a given level facilitates processing at that level. The results also support the hypothesis that the high frequency of the word and, or its role as a function word, or both, facilitate processing at levels higher than the letter.

Context Effects. Word frequency is not the only variable of importance, however. Manipulations of the surrounding word context also prove to be critical. More detection errors on the word and were made by subjects reading the scrambled-word and passage when the word and occurred between two like parts of speech, or in "appropriate context," than when it occurred between unlike parts of speech, or in "inappropriate context" [$t(23) = 3.12$, $p < .01$]. The mean number of errors on the word and was 3.29 for ands in appropriate contexts and 2.46 for ands in inappropriate contexts. These results suggest that when the word and occurs as part of a syntactically correct word group, it may be processed as part of a phrase-level unit.

List Passage. The processing of phrase-level units should be disrupted or eliminated by the use of the list passage, which not only removes the words from their adjoining spatial positions, but also alters the usual left-to-right reading pattern. As expected, the conditional error percentage for the list passage was at chance level and was significantly below that observed for the scrambled-word and passage [$t(31) = 5.37$, $p < .001$].

EXPERIMENT V

In Experiment V, we independently manipulate the level of the target and the highest level of processing available in the search passage. As in Experiment IV, scrambled-word passages in which phrase-level units are available are compared to list passages in which the processing of phrase-level units is impaired and the highest available processing level is the level of the word. Letter groups and and ant, both alone and embedded in other words, are used as targets in the first (Trigram Search) experimental condition, and the words and and ant are used as targets in the second (Word Search) condition.

According to our proposed set of hypotheses, detection of targets both at the letter group level and at the word level should be impaired by processing at the phrase level. Consequently, we should expect to continue to find more errors on the word and for the scrambled-word passage than for the list passage under both experimental conditions. Further, since we predict that the word ant in the scrambled-word passage is less likely than the word and to enter into phrase-level units, whether by virtue of its low frequency, its role as a content word, or the "inappropriateness" of its contexts in the present passage, we should expect to find substantially fewer errors on the word ant than on the word and in the scrambled-word passages. Moreover, we should expect to find no differences in detection errors on the word ant between the ant scrambled-word passage and the ant list, since phrase-level units are unlikely to be formed in either case.

We also investigate two plausible alternative hypotheses that could account for the preponderance of detection errors on the word the that was observed for subjects instructed to search for the the trigram. First, the word the is shorter, and therefore may be less conspicuous, than the necessarily longer target words containing the embedded the trigram. If the word and, like the word the, were missed on account of its length, there should be no difference in the pattern of errors between the scrambled-word passage containing the target *and* and the analogous passage containing the target *ant*, since the two targets are of equal length. The disproportionate number of detection errors on the word *the* in the previous experiments could also be explained by postulating that a letter-group target might be easier to locate when it is embedded within a word than when it constitutes an entire word. Under these assumptions, we would expect to find a different pattern of results when subjects search for targets at the word level than when they search for targets at the letter-group level. In particular, we would not expect to find a difference in the frequency of detection errors between *and* and *ant* word targets. As we shall demonstrate below, these two hypotheses seem to be ruled out by the present results.

Method

Subjects. The present experiment was conducted at the Mathematical Psychology Laboratory of the Rockefeller University. Forty-eight male and female young adults recruited from a newspaper advertisement served as subjects in the two experimental conditions: twenty-four subjects served in the Trigram Search condition, and twenty-four subjects served in the Word Search condition.

Design and Materials. We employed four passages of scrambled words. Three of the passages--the scrambled-word *and* passage, the scrambled-word *ant* passage, and the *and* list--were the same as those used in Experiment IV. The fourth passage was the *ant* list, which consisted of the same words as the *ant* passage, typed in five vertical columns. The order of the words and the type-case of the letters were the same as in the *ant* passage, but all punctuation marks were removed. The *ant* list was thus comparable to the *and* list.

Procedure. Subjects in the Trigram Search condition were asked to circle the letter group *and*, both alone and embedded in other words, in the *and* passage and list, and to circle the letter group *ant*, alone and embedded in other words, in the *ant* passage and list. Subjects in the Word Search condition were instructed to search for the words *and* and *ant*. As in the previous experiments, the subjects were told to read each passage at their normal reading speed and to read each list of words from top to bottom. The orders of presentation of *and* versus *ant* circling tasks and of the two passages within each task were counterbalanced across subjects. The reading time for each passage was determined with a stopwatch by the experimenter.

Results

Trigram Search. The results of the Trigram Search condition are summarized in Table 5, which includes for the four passages the means and the standard errors of the means for the total number of errors, the number of errors in "*and*/*ant* locations," the conditional percentage of errors in "*and*/*ant* locations" given an error, as well as the reading times in seconds. An error in an *and*/*ant* location is an error on the word *and* in the *and* passage and list, and in the corresponding locations on the word *ant* in the *ant* passage and list.

TABLE 5: Means and standard errors of means (in parentheses) for error frequencies, conditional percentages and reading times for passages with *and* or *ant* targets in the trigram search condition of Experiment V.

Target	Passage	Total errors	Errors in *and*/*ant* locations	Errors at *and*/*ant* locations given error %	N'	Reading times (in seconds)
and	scrambled word	7.54	4.66	69	22	58.3
		(1.31)	(.68)	(5)		2.7
	list	2.21	.83	37	15	58.1
		(.44)	(.23)	(8)		2.5
ant	scrambled word	3.37	.79	16	18	58.9
		(.73)	(.34)	(4)		3.1
	list	1.79	.29	20	13	58.4
		(.51)	(.12)	(8)		2.9

Note: Total N = 24.

There were no significant effects involving reading times. The mean conditional error percentage was significantly above chance for the scrambled-word *and* passage [$t(21) = 6.30$, $p < .001$], and was at chance level for the *and* list. The difference in error percentages between the scrambled-word *and* passage and the *and* list was significant [$t(13) = 2.55$, $p < .05$], whereas that between the corresponding scrambled-word *ant* passage and list was not [$t(10) = .71$, $p > .10$]. These results are consistent with our previous findings for the word *the* and provide further evidence that subjects may process frequent words in the scrambled-word passage in terms of phrase-level units. The increased accuracy on the list passage cannot be attributed to an increase in processing time on that passage, since reading times were equal for the two passages. Instead, a difference in reading strategy is implicated.

Further results rule out the possibility that subjects miss the word *and* solely on account of its length: the conditional error percentage for the scrambled-word *ant* passage was significantly less than that of the scrambled-word *and* passage [$t(38) = 7.41$, $p < .001$] and was significantly below chance level [$t(17) = 4.02$, $p < .01$].

Word Search. The results of the Word Search condition are summarized in Table 6, which includes for the four passages the means and the standard errors of the means for the total number of errors and for the reading times in seconds. Since the subjects were asked to search for the words *and* and *ant* only, the conditional error percentages cannot be derived for the present data.

TABLE 6: Means and standard errors of means (in parentheses) for error frequencies and reading times for passages with *and* or *ant* targets in the word search condition of Experiment V.

Target	Passage	Total errors	Reading times (in seconds)
and	scrambled word	2.00 (.44)	38.4 (3.2)
	list	.39 (.15)	31.2 (2.7)
ant	scrambled word	.41 (.15)	33.8 (3.0)
	list	.21 (.08)	30.2 (2.4)

Note: Total N = 24.

The subjects made more errors on the scrambled-word *and* passage than on any other passage. The difference between the *and* passage and the *ant* passage, in which every occurrence of the word *and* had been replaced by the less frequent word *ant*, was significant [$t(23) = 4.06$, $p < .001$], as was the difference between the *and* passage and the *and* list [$t(22) = 3.40$, $p < .01$]. In contrast, the difference between the *ant* passage and the *ant* list was not significant [$t(22) = 1.09$, $p > .10$]. More errors in the scrambled-word *and* passage were made when the word *and* appeared in its usual context, between two similar parts of speech, than when it occurred out of context [$t(23) = 2.62$, $p < .05$]. The mean number of errors on the word *and* was 1.25 for *ands* in appropriate contexts and 0.75 for *ands* in inappropriate contexts.

No difference was found between the *and* and *ant* lists, $t(22) = 0.94$, $p > .10$. Our hypotheses specify that word frequency will facilitate processing of a given word. On that basis we might have expected to find a difference between errors on the *and* and *ant* lists. However, since the level of the target in the Word Search condition and the highest level permitted by the list passage are both the level of the word, we should expect subjects in that condition to process all targets up to the word level. According to this line of reasoning, few errors are anticipated and any variation in the pattern of errors is due to random factors.

No significant differences in reading times were observed between the *and* passage and the *and* list or between the *ant* passage and the *ant* list. The reading times were in fact shorter for the two lists than for the two passages, though not significantly so.

These data show that subjects make a large number of errors on the word *and* relative to the less frequent word *ant* in scrambled-word passages, even when they are directed to search only for the word itself and not for any of the embedded trigrams. The fact that more ands were missed when the word and occurred in an appropriate context suggests that word context as well as word function or frequency play a major role in this phenomenon.

DISCUSSION

Subjects searching for instances of a given target letter in printed text make substantially more errors on the words the and and than on words containing embedded the and and trigrams (for example, thesis, handle). This effect is not due to word length, or to the pronounceability or the location of the target letter within the word. The contribution of the global word features of the frequent function words the and and is also unlikely. Rather, we show that the frequency or the function of the target words may be critical (Experiment IV).

Healy (1976) interpreted similar data by postulating that highly frequent words such as the may be read in terms of units larger than the letter. We now propose that under some circumstances, highly frequent word sequences including the words the and and may be read in terms of units larger than the word. Specifically, we propose that text may be processed at various levels, each of which involves units of a specific size, and that processing at these various levels occurs in parallel. We further propose that successful completion of processing at a higher level will terminate processing at all lower levels.

Evidence for these hypotheses comes from the observation that a disproportionate number of detection errors occur on the words the and and when subjects are searching for a given target letter and persist even when subjects are searching for the entire trigram (Experiments I, II, III and V). Three alternative explanations of this effect can be ruled out:

(1) Subjects fail to scan the words the and and because of their high predictability based on prior context (Corcoran, 1966). This alternative is ruled out by the finding that alterations in type-case and passage format reduce detection errors on the words the and and, although these alterations should not influence the contextual redundancy of those words (Experiments III and V). Furthermore, the similar notion that a combination of prior context and global features is sufficient to alert the subject to the presence of a function word, which is consequently not fully scanned (for example, Hochberg, 1970), can be eliminated by the finding that virtually no errors are made on the word ant whenever it replaces and in similar passages and in identical contexts (Experiment V), even though and and ant have very similar global features.

(2) Subjects make errors on the words the and and because of their short length, which makes them less conspicuous than the longer words containing the embedded letter strings. This second alternative is ruled out by the finding that subjects do not make a disproportionate number of detection errors on the less frequent word ant (Experiment V) even though and and ant are of equivalent lengths.

(3) Subjects make errors on the words the and and when searching for trigrams because a trigram is easier to locate when it is embedded within a word than when it constitutes an entire word. This third alternative is ruled out by the finding that subjects continue to make a large number of errors on the word and relative to the number on the word ant in scrambled-word passages even when they are instructed to search for the target words and or ant but not to respond to the and and ant letter groups embedded in longer words (Experiment V).

There are three manipulations that reduce the conditional percentages of detection errors down to, or close to, chance level, and each of these manipulations seems to be effective because it impairs phrase-level processing and hence disturbs the formation of reading units larger than the word. The most powerful manipulation is the change from the standard paragraph to a vertical list format. This procedure prohibits the natural left-to-right reading pattern and necessarily precludes the use of reading units larger than words (Experiments III, IV, and V).

The second manipulation, which does preserve the normal reading pattern, is the variation in context: the conditional percentages fall at chance level when syntactic units larger than the word are eliminated by placing the word the in syntactically inappropriate contexts (Experiment II). Similarly, more errors are made on the word the when it is followed by a noun (Experiments I and II) and on the word and when it is placed between two like parts of speech (Experiments IV and V). These results hold whether the subjects are instructed to search for a given target letter (Experiments I, II, and IV), trigram (Experiments I and II), or word (Experiment V).

The third manipulation, which involves changes in type-case, preserves both the normal reading pattern and the syntactic context but alters the perceptual global features of either the words or word groups. In particular, when passages are typed with every other letter (mixed letter) or every other word (mixed word) in capitals, the percentage of detection errors on the word the is reduced relative to that for passages with standard typing. The mixed-letter and mixed-word passages have similar effects; they both attentuate but do not reduce to chance levels the conditional percentages of detection errors on the word the.

Since all three manipulations disrupt processing at levels higher than the word, these results provide converging evidence which leads us to propose that familiar word sequences that often include the words the or and are read in terms of units larger than the word. We specifically suggest that these units are at the phrase level. In particular, they may be short syntactic phrases such as "boy and girl" or word frames such as "on the ___." Alternatively, the frequent function words, although separated from other words by word boundaries, may be read as prefixes or suffixes of the neighboring word. In any case, there is no evidence for processing of units much greater in size than the phrase, say on the order of the clause or sentence, since the preponderance of errors on the word the is no larger in a prose passage than in a standard scrambled-word passage [Experiment I and Healy (1976)].

It should also be made clear that the phrase-level units in question must be of high frequency. Otherwise, it would be difficult to explain why embedding the word the in a word unit such as weather leads to few detection errors, whereas embedding it in a phrase unit such as "on the ___" leads to many detection errors. These two types of situations differ in two respects. There is a difference between the levels of the units in question (word vs. phrase), and there is a difference between the frequency of the units in question (low for words vs. high for phrases). According to the hypotheses outlined above, it is the frequency of the given unit rather than its level that is critical in determining the likelihood of detecting embedded lower-level units.

As Healy (1976) has remarked, it is important to keep in mind that the reading units in question may be one of two possible types--perceptual units or response units. The perceptual units would be visual and the response units would be acoustic, presumably formed by phonetic recoding. The involvement of response units is suggested by the possibility that subjects searching for a target scan an acoustic image of the text rather than a visual image. This possibility seems reasonable on the basis of the evidence by Corcoran (1966) that subjects searching for the letter e are likely to miss those instances that are not pronounced, and evidence by Krueger (1970) that acoustic confusability retards letter detection. On the other hand, the effective manipulations of perceptual variables in the present study (Experiment III) suggest the involvement of visual units. At present we are, therefore, unable to choose between these two classes of units.

The present results, which are consistent with our set of hypotheses, are less clearly consistent with other models of the reading process. Serial processing theories are incompatible with the finding that subjects fail to detect a target at a given level when higher level units become available. Contextual redundancy theories are ruled out by the data from Experiments III and V (see above). Furthermore, explanations based on the notion of a speed-accuracy tradeoff cannot account for our observed results because our measure of conditional error percentages is independent of the subjects' absolute accuracy levels on this task.

It should be emphasized that we should not expect the particular results we obtained with the common function words the and and to generalize to less frequent words. Indeed, we have shown that the results for and do not apply to ant. The special properties of the frequent function words are what make them especially likely to be joined to other words in reading. However, although the specific results for the and and may not generalize to other words, the results for these words do throw light on how subjects read other words as well. In particular, the identity of the words surrounding the function words proves to be critical. When the function words are surrounded by appropriate neighbors, and only then, does a preponderance of detection errors occur on the function words. These results suggest that in appropriate syntactic contexts, neighboring words are read in conjunction with function words. Thus, whereas the earlier results of Healy (1976) demonstrated the critical role of the frequency of a given word in determining the occurrence of target detection errors, our present results demonstrate the critical role of the familiarity of a given word sequence.

In conclusion, although the relevance of the detection task to normal reading may be questioned, we argue that the present detection paradigm approaches the normal reading situation more closely than do many of the other letter detection paradigms in the literature (for example, Wheeler, 1970; Johnson, 1975). The occurrence of the same pattern of results when subjects are forced to process semantic characteristics of the words (Experiment III) as when the task does not specifically make such demands, strongly suggests that subjects may typically perform the detection task using processes employed in normal reading for meaning.

APPENDIX

Prose Passage of Experiment I

All week the weather was amazing. Even flowers in the park withered and became leathery under the sun's thermal rays. Children wearing no clothes bathed near the southern shore of the lake, while their mothers discussed other problems of psychotherapy and anesthesia. Panthers in the zoo surveyed the scene in a fatherly manner. As shadows lengthened, the air became ethereal and clouds began gathering on the horizon. 'Bother,' mumbled Alice, who was one of the sunbathers, 'I've hardly begun my thesis on the theory of medieval atheism, and I'd rather go and buy the earthenware jug I saw on Friday.'

REFERENCES

Corcoran, D. W. J. (1966) An acoustic factor in letter cancellation. *Nature* *210*, 658.

Doehring, D. G. (1976) Acquisition of rapid reading responses. *Monograph of the Society for Research in Child Development* *41*.

Estes, W. K. (1975) The locus of inferential and perceptual processes in letter identification. *Journal of Experimental Psychology: General* *104*, 122-145.

Estes, W. K. (in press) On the interaction of perception and memory in reading. In *Basic Processes in Reading: Perception and Comprehension*, ed. by D. LaBerge and S. J. Samuels. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Fisher, D. F. (1975) Reading and visual search. *Memory & Cognition* *3*, 188-196.

Gibson, E. J. (1965) Learning to read. *Science* *148*, 1066-1072.

Gibson, E. J. (1971) Perceptual learning and the theory of word perception. *Cognitive Psychology* *2*, 351-368.

Healy, A. F. (1976) Detection errors on the word *the*: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance* *2*, 235-242.

Hochberg, J. (1970) Components of literacy: Speculations and exploratory research. In *Basic Sdies on Reading*, ed. by H. Levin and J. P. Williams. (New York: Basic Books).

Johnson, N. F. (1975) On the function of letters in word identification: Some data and a preliminary model. *Journal of Verbal Learning and Verbal Behavior* *14*, 17-29.

Krueger, L. E. (1970) The effect of acoustic confusability on visual search. *The American Journal of Psychology* *83*, 389-400.

Kučera, H. and W. N. Francis. (1967) *Computational Analysis of Present-Day American English*. (Providence: Brown University Press).

LaBerge, D. and S. J. Samuels. (1974) Toward a theory of automatic information processing. *Cognitive Psychology* *6*, 293-323.

McClelland, J. L. (1976) Preliminary letter identification in the perception of words and nonwords. Journal of Experimental Psychology: Human Perception and Performance 2, 80-91.

Spoehr, K. T. and E. E. Smith. (1973) The role of syllables in perceptual processing. Cognitive Psychology 5, 71-89.

Underwood, B. J. and R. W. Schulz. (1960) Meaningfulness and Verbal Learning. (Philadelphia: Lippincott).

Wheeler, D. D. (1970) Processes in word recognition. Cognitive Psychology 1, 59-85.

Some Observations on the Perception of [s]+Stop Clusters

Peter J. Bailey[†] and Quentin Summerfield[††]

ABSTRACT

A series of experiments is reported that investigated the pattern of acoustic information specifying place and manner of stop consonants in medial position after [s]. In both production and perception, information for stop place includes the spectrum of the fricative at offset, the duration of the silent closure interval, the spectral relationship between the frequency of the stop release burst and the following periodically excited formants, and the spectral and temporal characteristics of the first formant transition. Similarly, the information for stop manner includes the duration of silent closure, the frequency of the first formant at the release, the magnitude of the first formant transition, and the proximity of the second and third formants at release. A relationship was shown to exist in perception between the spectral characteristics of the first formant and the duration of the silent closure required to hear a stop. This appears to reciprocate the covariation of these parameters in production across different places of articulation and different vocalic contexts. The existence of perceptual sensitivity to a wide range of the acoustic consequences of production questions the efficacy of accounts of speech perception in terms of the fractionation of the signal into elemental cues that are then integrated to yield a phonetic percept. It is argued that it is an error to ascribe a functional role in perception to such "cues", that only have reality in their operational role as physical parameters whose manipulation can change the phonetic interpretation of a signal. It is suggested that the metric of the information for phonetic perception cannot be that of the cues; rather, a metric should be sought in which acoustic and articulatory dynamics are isomorphic.

[†] Department of Psychology, University of York, U.K.

[††] MRC Institute of Hearing Research, The University of Nottingham, U.K.

Acknowledgment: This work was carried out while the authors were supported by N.A.T.O. post-doctoral research fellowships from the U.K. Science Research Council. We should like to thank Alvin Liberman for his hospitality and encouragement, and Bruno Repp, Francis Ganong and Michael Studdert-Kennedy for comments on an earlier draft of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 2 (1978)]

INTRODUCTION

If one hundred milliseconds of silence are inserted between the fricative and vocalic portions of the word "slit," the word is heard as "split"; a similar operation on the word "sag" leads to the perception of "stag" (Bastian, 1959). It is generally concluded that silence, or at least a period of little acoustic energy, is a cue to stop consonant manner (Raphael, Dorman and Liberman, 1976). The initial intention of the present experiments was to specify the acoustic information for stop place; that is, to determine why, in the examples above, for instance, [p] is heard in "split" and [t] in "stag". A partial answer may be found in the data reported by Lotz, Abramson, Gerstman, Ingemann and Nemser (1960) and by Reeds and Wang (1961). These authors demonstrated that, when the fricative segment is removed from prevocalic [s]+stop clusters, the place information for the stop is preserved in the vocalic segment: removing [s] from "spill" and "score," for instance, leaves "bill" and "gore," respectively. Analogously, removing [s] from a natural production of "split" leaves "blit." However, when [s] is removed from "slit" no stop is heard: the vocalic portion sounds like "lit." What are the properties of this segment that give rise to a bilabial percept?

In a pilot experiment we used a serial resonance synthesizer to create a set of ten steady-state vowels based on the formant frequency data for adult males published by Peterson and Barney (1952). Each vowel was preceded by a period of [s] friction and 100 msec of silence. Some of these syllables were identified as [s]+vowel, others as [s]+stop+vowel. Overall, the probability of hearing a stop was inversely related to the frequency of the first formant, while the place of production of the stop appeared to be determined by the frequency of the second formant, approximately in accordance with the principle of formant loci (Delattre, Liberman and Cooper, 1955). In general, [s]+[u] was heard as [spu], [s]+[e] as [ste], and [s]+[i] as [ski]. No stops were heard in syllables incorporating vowels with high first formants; [s]+[a] gave [sa], for example. Stop percepts were strongest with [spu] and [ski], while those with [ste] were weak. This contrasts with the results of Delattre et al. (1955), who noted that in two-formant consonant vowel (CV) syllables, the most compelling percept that could be obtained with a steady second formant was of an initial alveolar stop. However, in our pilot experiment, the effects attributed to the second formant were confounded not only with the frequency of the first formant, but also with variation in the relative levels of the formants across the set of vowels as a result of the serial connection of the synthesizer resonances. In the first two experiments reported below, we controlled these variables by using parallel resonance synthesis to examine the effects of second formant frequency on place of production of stops perceived in [s]+silence+vowel syllables.

EXPERIMENT I

Stimuli

A continuum of 12 two-formant steady-state vowels was prepared with the parallel resonance synthesizer at the Haskins Laboratories. Throughout the continuum the first formant frequency was fixed at 260 Hz, while the second formant frequency increased from 616 Hz to 2307 Hz in steps of approximately 154 Hz. The intensities of the first and second formants were the same[1]. The amplitude rise and fall times for the vowels were 75 msec and 50 msec, respectively. Each vowel was 250 msec in duration. The end-points of the continuum were acceptable examples of the English vowels [u] and [i]. The midrange stimuli, on the other hand, were not English vowels but approximated the central vowels [ʉ] and [ɨ], as found, for example, in Swedish and Russian. A steady-state [s] of 120 msec duration was created with an OVE IIIc serial resonance synthesizer[2]. The fricative formants were set to 3489 Hz and 4532 Hz, and the fricative antiresonance was set to eliminate energy below 2000 Hz. The amplitude rise-time of the [s] was 40 msec and its fall-time was about 10 msec. The twelve vowels and this [s] were low-pass filtered at 9.7 kHz and digitized with a sampling rate of 20 kHz. Two stimuli were prepared from each vowel by appending it to the [s] with either 10 msec or 100 msec of silence intervening. A randomization of 240 trials was recorded in which each of the 24 stimuli appeared ten times. The intertrial interval was three seconds and the trials were arranged in blocks of ten, with an extra three second pause between blocks.

Subjects and Procedure

Fourteen undergraduate volunteers served as subjects. They had normal hearing in both ears and learned English as their first language in the U.S.A. The two authors (who learned English in the U.K.) also took part in the experiment.

Stimuli were presented binaurally through Grason-Stadler TDH-39 headphones at a constant peak listening level of 75 dB SPL. Subjects were instructed that they would hear a randomization of [s]+vowel and [s]+stop+vowel syllables. Their task was to write down the letter 'V' if they heard [s]+vowel, and one of the letters 'P','T' or 'K' if they heard either [sp]+vowel, [st]+vowel or [sk]+vowel. They were also permitted the response 'Q' to indicate a stop percept without identifiable place of production, and the response 'O' to indicate a consonant other than [p], [t] or [k]. Subjects listened first to a 24-trial practice sequence that included one exemplar of each stimulus, and then to the 240-trial test sequence.

---

[1] Nominal half-power bandwidths for the first and second formants were 60 Hz and 90 Hz respectively.

[2] In all the experiments reported here, the OVE synthesizer was used to create frication, as its fricative branch permits control over two poles and a simulated zero, in contrast to the single pole available in the Haskins parallel resonance synthesizer.

Results

The data from the 14 volunteer subjects and those from the two authors did not differ qualitatively and were pooled together. Each of the 12 stimuli with 10 msec of silence was identified as [s]+vowel on at least 93 percent of its presentations, whereas no stimulus with 100 msec of silence was identified as [s]+vowel on more than 10 percent of its presentations. The comparison emphasizes that the concatenation of [s]+silence+vowel is not, _per se_, sufficient for the percept of a stop: a critical amount of silence is required. One rationalization of the result would be that the vocalic segments could themselves be heard as CV syllables and that a critical duration of silence is required for the initial consonant to evade masking by the [s]. This possibility was tested in a subsidiary experiment. A randomization was recorded that included six instances of each of the twelve vowels and six instances of twelve stimuli derived from these vowels by reducing their amplitude rise-time from 75 msec to 5 msec. The randomization was presented to six experienced listeners who were instructed to identify the vowel and to indicate whether or not it was preceded by a stop. Out of 864 identifications there were only four reports of an initial stop. These occurred for stimuli with the more rapid rise in amplitude. We conclude that the vowels used in Experiment I did not, of themselves, predispose the percept of a stop.

The results from the twelve stimuli with 100 msec of silence are displayed in Figure 1 where the percentages of 'V', 'P', 'T' and 'K' responses are plotted against the stimulus number, and hence the frequency of the second formant in the vowel. For the sake of clarity, responses in the 'Q' and 'O' categories have been combined. There was a slight tendency for 'O' responses to predominate at low second-formant frequencies, and for 'Q' responses to predominate at high second-formant frequencies. Variation in the frequency of the second formant produced systematic effects upon the place of production of the perceived stop. The probability of [p] percepts decreased and the probability of [k] percepts increased as the second-formant frequency was raised. The fact that [p] predominated when the second formant was below 1695 Hz is consistent with the perception of [p] in the sequence [s]+silence+[lit], where in the vocalic segment the second formant typically onsets below 1000 Hz (Fant, 1960).

As in the pilot experiment, the vowels [u] and [i] gave reasonably strong percepts of [spu] and [ski], respectively. The [t] category was almost absent here, suggesting that its weakness in the pilot experiment did not result either from an insufficiently low frequency of the first formant for a stop to be heard, or from a relatively low level of second formant intensity, in mid-central vowels. As we have already noted, the locus principle would predict that [t] would be used as a response to the stimuli numbered 8, 9 and 10, whose second formants span the 1800 Hz alveolar locus. In fact, [t] was made as a response to these stimuli on only 3 percent of their presentations, while the largest proportion of [t] responses made to any stimulus (10 percent) was made to stimulus number 5 in which the second formant was set to 1232 Hz. The principle of formant loci appears to be insufficient to explain the perception of place of articulation in these stimuli where the stops were not in absolute initial position.

One reason for the presence of [t] percepts in the pilot experiment and their absence here could be that the vocalic portions of the pilot stimuli incorporated five formants and thus included energy above 2.5 kHz, whereas those in Experiment I, incorporating only two formants, had no energy above 2.5 kHz. This would be consistent with the proven perceptual efficacy of energy around 3 kHz for specifying the release of an alveolar stop (Fant, 1960). Accordingly, Experiment I was repeated using stimuli whose vocalic portions included a third formant fixed at 3026 Hz.

EXPERIMENT II

Procedure

The stimuli in Experiment II were derived from those in Experiment I by adding a fixed third formant at 3026 Hz to the vocalic portion of each stimulus. The level of the third formant was -15 dB with respect to the first and second formants, and its nominal half-power bandwidth was 120 Hz. Identification tapes were prepared as before and administered to eight volunteer subjects and to the two authors. The instructions and response alternatives were the same as in Experiment I.

Results

The stimuli with 10 msec of silence were not identified as [s]+vowel as consistently as their counterparts in Experiment I. On average they were identified in this way on only 86 percent of their presentations, due largely to a tendency for stimuli with high second formant frequency to be identified as [sk]+vowel. This result presages that obtained with the stimuli incorporating 100 msec of silence whose identification functions are shown in Figure 2. Data from ten listeners, including the two authors, are displayed. The only systematic difference between the results of Experiments I and II was an increase in the proportion of [k] responses to stimuli with high second formants in the second experiment. The incorporation of a fixed third formant in the vocalic portions of these stimuli strengthened the velar category, but, although the overall proportion of [t] responses increased from 3 percent to 6 percent, no clearly defined [t] category emerged.

Discussion

On the basis of perceptual experiments with two-formant patterns, Delattre et al. (1955) deduced fixed second formant loci of 720 Hz for bilabial stops, 1800 Hz for alveolars and 3000 Hz for velars before front vowels; a locus at a lower and more variable frequency was found for velars before back vowels. Stevens and House (1956) largely confirmed these results using an electrical analogue of a vocal tract, but showed that neither the bilabial nor the velar loci are completely fixed. Both loci vary as a function of the extent to which the arrangement of the articulators during closure anticipates their organization for the following vowel. On the assumption of maximum anticipatory co-articulation, the locus for [b] can range from 700 Hz to 1500 Hz, and that for [g] from 600 Hz to 2500 Hz. We have already noted the apparent anomaly between our failure to find alveolar percepts in Experiments I and II, and the observation of Delattre et al. (1955) that

"... of the three stops that are produced when the straight second formants are at the loci the [d] (at 1800 Hz) is the most compelling." (p.771). In their stimuli, the information for stop manner was embodied in a rising first formant transition; we have to explain why alveolar percepts are absent here when information for stop manner is embodied in a period of silence. The Stevens and House (1956) simulation demonstrated that the loci were not arbitrary perceptual constructs; they are a direct consequence of stop consonant articulation. The ranges of formant loci computed by Stevens and House could provide a basis for rationalizing the distributions of [p] and [k] responses in Experiments I and II. However, a reliance on formant loci as an explanatory principle would still require an *ad hoc* account of the absence of the [t] category.

The optimal strategy for explaining the perception of an articulatory event would be to consider not just one but all of the acoustic consequences of that event. The present data should be rationalized by determining which natural production is represented most faithfully by our, albeit schematic, stimuli. These stimuli may be characterized as follows: first, there are no formant transitions in the vocalic portions; second, there are no release bursts; third, as a consequence, the spectral energy at and following release is contiguous and continuous. The property of spectral contiguity was identified by Fant (1973) as a general characteristic of initial velar stops: "(for) [k][g] spectral energy is concentrated, strong and continuously connected, without rapid initial transitions to the formant carrying the main pitch of the vowel..." (p.135). As Fant's spectrographic reference material shows, the absence of initial transitions in velars is strictly characteristic only of front vowels when the second formant frequency is high; appreciable second formant transitions are evinced before back vowels where the second formant is at a lower frequency. For bilabial stops, Fant (1973) notes that: "...spectral energy is weak, more spread than in [k][g], and with an emphasis on a lower frequency than [t][d]. Initial transitions are rapid and rising" (p. 135). Thus, in contrast to velar stops, transitions characterizing initial bilabials are minimal before vowels with low second formants, and increase in magnitude before vowels with higher second formants.

Taking these observations in conjunction with the ranges of second formant loci reported by Stevens and House (1956), we can account for the pattern of results in Experiments I and II. Bilabial percepts would be expected (in the absence of a prominent burst), when the second formant onsets at a low frequency with no formant transition. Velar percepts, on the other hand, would be expected when the second formant onsets at a high frequency without any formant transition. Inevitably, there is spectral contiguity between energy at the release and that in the following vowel, although there is no release burst as such. The larger number of [k] responses in Experiment II following the introduction of a fixed third formant probably resulted both from the presence of higher frequencies in the vocalic onset, and from the proximity of the second and third formants when the second-formant frequency was high, given that proximity of the second and third formants at onset is a characteristic of the production of velar stops (Fant, 1960; Stevens, 1975). The absence of alveolar percepts in our data is consistent with Fant's (1973) observation that for [t] and [d] "spectral energy is spread, generally strong, with emphasis on higher frequencies than in [p] and [b] and extending higher than the main [k][g] formant" (p. 135). As this quote implies, and as Fant's spectrograms show, spectral discontinuity between the release and the periodically excited formants characterizes alveolar stops before all vowels. This discontinuity accompanies the sudden increase in the length of the front cavity as the major constriction switches from the alveolar region to a more dorsal position characterizing the vowel. We suspect that the lack of any spectral discontinuity in our stimuli at release is a major contributor to the absence of alveolar percepts. These observations appear to provide a coherent account of our data. The differences between our results and those of Delattre et al. (1955), where manner information was carried by formant transitions, can be understood in terms of the reciprocal relationship between the importance of bursts and transitions in the perception of stop place (Cooper, Delattre, Liberman, Borst and Gerstman, 1952; Dorman, Studdert-Kennedy and Raphael, 1977).

The major claims of the foregoing account were tested below in Experiment III by using stimuli that included a prevocalic release burst. We expected to find that the place of articulation of stops perceived in such stimuli would be determined both by the frequency of the burst, and by the relationship between burst frequency and the frequency of the second formant at onset. Specifically, it was predicted that [p] percepts would be obtained infrequently with a concentrated burst but might appear when the burst was at a lower frequency than the second formant; [k] percepts were predicted when the burst frequency and the second formant onset frequency were contiguous; [t] percepts were predicted when the burst frequency was higher than and not contiguous with the onset frequency of the second formant.

EXPERIMENT III

Stimuli

Twenty-five stimuli were constructed as before by combining a fricative segment synthesized using the serial resonance synthesizer, with a vocalic segment synthesized using the parallel resonance synthesizer. The fricative segments consisted of an [s] of 120 msec duration followed by 100 msec of silence and, in this experiment, a 25-msec release burst. The [s] was spectrally identical to the [s] segments used in Experiments I and II. Five bursts were created by setting the lower fricative pole to 1509 Hz, 1957 Hz, 2466 Hz, 3019 Hz and 3489 Hz, respectively; in each case the higher fricative pole was set to 6 kHz and was cancelled with the antiformant. Bursts synthesized in this way increase in intensity as their frequency rises. Five two-formant steady-state vowels were synthesized. Each was 250 msec in duration, with the first formant set to 260 Hz. The second formant was set to 1386 Hz, 1772 Hz, 2156 Hz, 2540 Hz and 2910 Hz, respectively in the five patterns. The levels of the two formants were the same and their amplitude rise-time was 20 msec. The five fricative segments and the five vowels were low-pass filtered at 4.9 kHz and digitized with a sampling rate of 10 kHz. Twenty-five test stimuli were constructed by preceding each of the vocalic segments with each of the fricative segments. Thus, each stimulus consisted of the sequence [s] (120 msec) + silence (100 msec) + burst (25 msec) + vowel (250 msec). A 25-trial practice sequence consisting of a single randomization of these 25 stimuli, and a 250-trial test sequence consisting of ten concatenated randomizations were recorded. The intertrial interval was three seconds.

Subjects and Procedure

The two authors and six experienced listeners who were unaware of the structure of the stimuli listened first to the practice sequence and then to the test sequence. Stimuli were presented binaurally through Grason-Stadler TDH-39 headphones at a constant peak listening level of 75 dB SPL. The eight listeners were required to identify the stop heard in each syllable as either [p], [t] or [k] but to indicate with a question mark any response of which they were not confident.

Results and Discussion

Of the 2000 responses only ten indicated ambiguous percepts and these will not be distinguished from other responses. The data of the two authors did not differ systematically from those of the other six listeners. Figures 3a - 3e display the data pooled over all eight subjects. Each panel shows the percentage of [p], [t] and [k] responses made to the five stimuli with the same burst frequency. In each case the abscissa plots the frequency of the second formant in the vocalic portion of the stimuli. As predicted, there were few [p] responses. They appeared when the burst frequency was low and below the onset of the second formant; for example, in Figure 3a when F_2 was at 1772 Hz, in Figure 3b when F_2 was at 2540 Hz and in Figure 3c when F_2 was at 2910 Hz. The [k] percepts were most likely when the burst frequency was close to the onset frequency of the second formant, although an exception to this generalization is found in Figure 3a. Here, with the burst at its lowest frequency and hence its lowest intensity, a pattern of results most akin to that of Experiment I would be expected. However, a comparison of Figure 3a with Figure 1 shows that the presence of a burst at this frequency has had the effect of increasing the probability of [k] percepts. The proportion of [t] responses increased with both the burst frequency and the size of the frequency difference between the burst and the second formant onset. In each of Figures 3b, 3c and 3d, [t] percepts predominated when the second formant was low and [k] percepts predominated when the second formant was high. The crossover between alveolar and velar responses occurred at higher second formant onset frequencies as the frequency of the burst increased.

The results of Experiment III are consistent with earlier results (Liberman, Delattre and Cooper, 1952) and with our rationalization of the data from Experiments I and II. The [t] is perceived medially after [s] and before a vowel in the absence of periodically excited formant transitions, if a burst initiates the vowel with a center frequency at least 400 Hz above the main formant in the vowel. The appropriate complement to this result would be the demonstration that [st] percepts occur in the absence of a burst, provided that the vocalic portion of the stimulus incorporates periodically excited formant transitions. This was an ancillary finding of Experiments VI and VII.

The results of the first three experiments confirmed that a complete account of the perception of stop consonants in medial position after [s] depends upon an understanding of the acoustic consequences of the underlying articulatory event as a whole. So far, only the release phase of the event has been considered. Before examining the perceptual concomitants of the constriction and occlusion phases, we sought to determine how the acoustic properties of natural productions of [s]+stop+vowel sequences vary according to the place of production of the medial stop.

PRODUCTION DATA

Procedure

The two authors read a randomization of [sSV] syllables, where S was one of [p], [t] or [k], and V was one of [i], [ɜ], [a] or [u]. The syllables were uttered in the phrase "Now hear [sSV] please" at a natural rate cued by a visual metronome. Five tokens of each stop-vowel combination were recorded. The recordings were low-pass filtered at 4.9 kHz and digitized at a sampling rate of 10 kHz. Both the sampled waveform and a hardware spectrum analysis of successive 12.8 msec segments of the signal were displayed. The computer system allows a cursor to be aligned to measure the duration of any segment of the waveform to an accuracy of 0.2 msec. The spectral section of the 12.8 msec segment containing the cursor is also displayed. Two spectral measures were made in each token; they were the frequencies of the lowest peak in the spectral sections containing (a) the offset of the fricative portion and (b) the release of the stop.[3] Three intervals were measured; these were (a) the duration of the fricative portion, (b) the period of silent closure and (c) the period of aspiration following release prior to the first voicing pulse.

The spectral measurements are summarized in Table 1 where the average frequencies of the lowest spectral peak at closure and release are tabulated for each syllable. With the exception of the bilabial stops, whose burst frequencies could not be estimated reliably with our measurement procedure, the release burst spectra for medial stops following [s] are in reasonable agreement with measures from stops in initial position (for example, Fant, 1973), as required by the rationalization of the results of the first three experiments. Fricative offset spectra also show a consistent pattern: spectral peaks in the fricative offset are at lower frequencies in syllables with bilabial stops than with either alveolars or velars. This reflects the lengthening of the cavity in front of the fricative source as bilabial closure is made, and can sometimes be seen in spectrographic displays as a rapid downward transition at the end of the fricative. Given that different spectral changes are concomitants of stops articulated at different places, we should expect to find perceptual sensitivity to the spectral characteristics

---

[3] The hardware spectrum analysis is relatively coarse-grained in both frequency and time and permits only approximate estimates of the frequencies of the spectral peaks corresponding to any particular portion of the waveform.

Table 1: Spectral Measurements (kHz) lowest frequency peak.

	Speaker 1 : PJB		Speaker 2 : AQS	
Syllable	[s] offset	Release burst	[s] offset	Release burst
[spi]	3.1	*	3.6	*
[spa]	3.4	*	3.7	*
[spu]	3.1	*	3.8	*
[spɜ]	3.1	*	3.7	*
Mean	3.2		3.7	
[sti]	3.7	3.1	4.5	3.6
[sta]	3.9	3.4	4.2	3.8
[stu]	3.5	2.7	4.5	3.0
[stɜ]	3.7	3.1	4.6	3.3
Mean	3.7		4.4	
[ski]	3.8	3.0	4.2	2.9
[ska]	3.8	1.8	3.9	1.4
[sku]	3.7	1.6	3.9	1.4
[skɜ]	3.4	1.8	4.0	1.5
Mean	3.7		4.0	

* Spectral analysis too coarse to measure burst spectrum.

of the fricative offset in judgements of stop place in [s]+stop clusters. This was investigated below in Experiment IV.

The duration measurements are tabulated in Table 2. The means of the standard deviations of the durations of friction, silence and aspiration were 17.8 msec, 15.0 msec and 4.2 msec for speaker 1, and 10.8 msec, 8.1 msec and 5.2 msec for speaker 2. Insofar as they are comparable, these figures are in agreement with the variability in similar duration measurements reported elsewhere (for example, Klatt, 1974; 1975). There are three noteworthy aspects of these data. First, both speakers produced longer silent intervals during bilabial closure compared to alveolar or velar closure. Thus, closure duration appears to be another characteristic of place of production, and Experiments V and VI were designed to measure perceptual sensitivity to this variable. Second, both speakers produced shorter silent intervals before the vowel [a] compared to the other three vowels. To the extent that this difference reflects an inverse correlation in production between the openness of the vowel and the duration of the silent closure, we should expect to find a trading relationship between the magnitude of the first formant transition and the amount of silence required for the perception of a stop in an [s]+stop cluster. This was examined in Experiment VII. The third aspect of the data in Table 2 is parenthetic to our main interest here. It is that the total period of devoicing, that is, the sum of the durations of friction, silence and aspiration, is less variable across stop place and vowels than is any one of its component durations. We have noted a similar tendency for total durations of devoicing to be relatively invariant across place in productions of [bɜSV] and [bɜSrV], where S was one of [p], [t] or [k] and V was [i] or [a] (syllables such as [bɜpa] and [bɜkri], for instance). Both results suggest that, at any particular speaking rate, control in production (of stressed syllables) is exercised over the laryngeal event as a whole, and not over the temporal microstructure of the sequence of segments that are the acoustic consequences of that event.

It would appear that the spectral properties of stop release after [s] accord with our interpretations of the first three experiments. In addition, the production data have shown that concomitants of stop place, to which perceptual sensitivity may be demonstrable, exist in both the spectrum of friction offset and the duration of stop closure. They have also suggested a possible articulatory basis for a trading relationship between the duration of stop closure and the characteristics of the first formant at voicing onset in the perception of [s]+stop clusters. These possibilities are explored in the four experiments reported below.

EXPERIMENT IV

Experiment IV was designed to investigate the influence of variation in the spectral properties of fricative offset on the perception of stop place in [s]+silence+vowel syllables. Specifically, we sought to determine whether the relationship between the position of the [p]-[k] boundary and the frequency of F_2, shown in Figures 1 and 2, could be changed systematically by varying the spectral properties of the final 35 msec of the [s] frication.

Table 2: Duration measurements (msec).

	Speaker 1 : PJB				Speaker 2 : AQS			
		Durations of				Durations of		
Syllable	Friction	Silence	Aspiration	Total	Friction	Silence	Aspiration	Total
[spi]	174.9	130.7	14.8	320.4	166.5	100.4	12.2	279.1
[spa]	201.1	108.8	15.5	325.4	156.6	96.3	12.9	265.7
[spu]	185.9	142.7	15.6	344.2	160.6	101.1	13.0	274.7
[spɜ]	185.7	127.3	15.1	328.1	166.3	100.3	10.1	276.8
Mean	186.9	127.4	15.3	329.6	162.5	99.5	12.1	274.0
[sti]	212.1	93.1	33.0	338.2	181.4	67.0	21.2	269.6
[sta]	214.1	66.8	21.1	302.0	179.8	60.6	17.5	257.9
[stu]	222.3	100.2	31.0	353.5	171.3	75.9	20.2	267.4
[stɜ]	218.7	83.0	26.0	327.7	182.8	75.1	19.8	277.8
Mean	216.8	85.8	27.8	330.4	178.8	69.7	19.7	268.2
[ski]	203.6	86.9	50.9	341.4	186.6	82.9	25.6	295.1
[ska]	212.7	65.2	35.9	317.4	180.7	63.9	23.4	268.0
[sku]	216.0	82.2	44.5	342.7	181.7	87.0	26.6	295.3
[skɜ]	220.5	87.1	44.4	352.0	195.1	77.4	21.4	293.9
Mean	213.2	80.3	44.4	337.9	186.0	77.8	24.5	288.1
Overall Mean	205.6	97.8	29.0	332.8	175.8	82.3	18.7	276.8
SD	15.7	24.9	13.0	15.3	11.4	14.7	5.5	12.3
SD/Mean	0.076	0.255	0.448	0.046	0.065	0.179	0.294	0.044

Stimuli

Forty stimuli were created by combining each of four [s] segments with each of ten vocalic segments. One hundred msecs of silence intervened between the two types of segment. Both segments were created with the serial resonance synthesizer. Each [s] segment was 150 msec in duration with amplitude rise and fall times of 50 msec and 15 msec, respectively. Over their first 115 msec, the four [s]s were constant in frequency with the fricative formants set to 3917 Hz (K1) and 4932 Hz (K2). The antiformant was set to eliminate energy below 2000 Hz. Different patterns of spectral change distinguished the final 35 msec of the four fricatives. In pattern S1, the two fricative formants rose linearly to 4936 Hz and 6038 Hz. In pattern S2 they remained at their steady-state values of 3917 Hz and 4932 Hz. In pattern S3, the lower fricative formant fell to 3019 Hz. In pattern S4, the lower fricative formant fell to 1957 Hz. The rise time of the vocalic portions was 30 msec. Over this duration, F_1 rose linearly from 200 Hz to a steady-state value of 299 Hz. The third formant was constant at 3199 Hz. The ten vocalic segments were distinguished by the frequencies of their constant second formants that ranged from 600 Hz to 2400 Hz in steps of approximately 200 Hz. A practice sequence consisting of a single randomization of the 40 stimuli, and a test sequence consisting of five concatenated randomizations were recorded.

Subjects and Procedure

Eight undergraduates served as subjects. They were phonetically naive, possessed normal hearing in both ears, and learned English as their first language in the U.S.A. They were instructed to identify the medial stop in each syllable as either [p], [t] or [k]. They listened first to the practice sequence and then twice to the test sequence. In this way, 10 identifications of each syllable by each subject were collected.

Results

In Figure 4, the data from the eight listeners have been pooled and are displayed in four graphs, one for each of the fricative patterns whose spectral specifications are schematized in the inserts. The percentages of [p], [t] and [k] responses are plotted as a function of the stimulus number, and hence of the second formant frequency in the vocalic portion of the stimuli. There is only a minimal difference between the patterns of [p] and [k] identifications of stimuli with fricative S1 and those with fricative S2; the number of [t] responses, already small, decreased slightly with the frequency of the fricative at offset. This trend continued through patterns S2, S3 and S4 as the fricative offset frequency was further reduced and is significant when assessed in an analysis of variance ($F_{3,21} = 3.40$; $p < 0.025$). However, the main result of the experiment is that the proportion of [p] responses increased at the expense of [k] responses as the offset of the lower fricative formant was reduced between patterns S2, S3 and S4. Overall, this effect is significant ($F_{3,21} = 9.26$; $p < 0.01$). Planned comparisons between adjacent series show that the only significant difference in the proportion of [p] responses is that between S2 and S3 ($F_{1,21} = 9.55$; $p < 0.01$).

In natural productions of [s]+stop clusters, different spectral changes in the offset of the [s] accompany stop closure at different places in the vocal tract (see Table 1). The results of Experiment IV show that the perceived place of a stop is influenced by the spectral properties of the [s] immediately prior to stop closure. However, although consistent, the effect is small. It is manifest as an increase in the region of ambiguity between bilabial and velar responses, and not as an increase in the number of stimuli identified unequivocally as bilabial. Nevertheless, these data, taken together with those of the previous experiments, show that, just as the event of stop consonant production occurs over time, so the acoustic information that specifies the identity of a stop for a perceiver is distributed over time.

The experiments that have been described so far have demonstrated perceptual sensitivity to the spectral properties of the segments bounding the period of stop closure. The following two experiments examine the influence of the duration of stop closure itself on the perception of place of articulation of stops in [s]+stop clusters.

EXPERIMENT V

In this experiment the size of the silent interval reflecting stop closure was varied to create four series of syllables, each varying from [s]+vowel to [s]+stop+vowel. The four series were distinguished by different values of second-formant frequency in the vowel. These values were chosen on the basis of the results of Experiment I to give two [su-spu] series, a [si-ski] series and a series ambiguous between the two. We wished to determine whether the size of the silent interval necessary for the perception of a particular stop varied as a function of either the acoustic specification of the stimulus, or its phonetic interpretation, or of both these factors.

Stimuli

Four 11-member [s]+vowel to [s]+stop+vowel series were created by inserting an increasing duration of silence between the [s]-friction and the vowel. The two-formant vocalic segments were identical to those in stimuli 1, 4, 9 and 12 in Experiment I. Their first formants were set to 260 Hz. Their second formants were set to 616 Hz, 1075 Hz, 1845 Hz and 2307 Hz, respectively. The [s] friction was the same as that used in Experiment I. For a given series, the duration of interpolated silence ranged from 0 msec to 100 msec in 10 msec steps. Two sequences were recorded for identification. One was a 44-trial practice sequence, the other was a 440-trial test sequence containing ten instances of each of the 44 stimuli.

Subjects and Procedure

Fifteen subjects, with the same qualifications as those who served in Experiment IV, listened first to the practice sequence and then to the test sequence. The stimuli were presented under the same conditions as in the earlier experiments. Subjects were instructed to identify each stimulus as either [s]+vowel or [s]+stop+vowel. In addition to a response for [s]+vowel, four response alternatives were provided for percepts of [p], [t], [k] and a glottal stop; a fifth alternative was provided for any stop not in these categories.

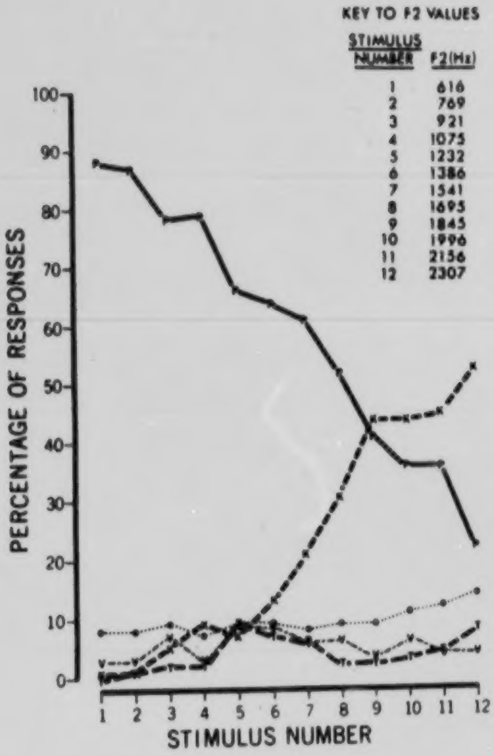
Results and Discussion

Figures 5a to 5d display the data corresponding to each test series pooled over fifteen subjects. Each graph plots, as a function of the duration of the silent interval, the percentage of [s]+vowel responses (V), [s]+stop+vowel responses (S), and the breakdown of the stop category into individual functions for [p], [t] and [k], glottal and other stop responses (Q). Functions are not shown for response categories that received fewer than 10 percent of the total number of responses.

The production data in Table 2 show that at any given rate of speech, stop closures for bilabials are typically longer than are those for alveolars and velars. The present experiment was designed to determine whether velar stops are perceived with shorter silent intervals than bilabials. We can contrast two extreme hypotheses. One, an "acoustic" hypothesis, suggests that as the frequency of the second formant rises, the probability of hearing any stop at a particular silent interval increases. Since Experiments I and II have shown that the frequency of the second formant also determines stop place, bilabials would be heard with longer stop closures than velars. A different hypothesis would be that a phonetic decision about stop place determines the criterial duration of silence characterizing stops with that place of articulation. A comparison of the four graphs in Figure 5 appears to lend some support to the acoustic hypothesis. The crossover between [s]+vowel and [s]+stop+vowel responses in the pooled data occurred at shorter durations of silence as the frequency of the second formant in the vowel was raised. However, this trend does not appear consistently in the data of individual subjects, either when represented as percentages of [s]+vowel responses, or when represented as 50 percent crossovers on the [s]+vowel identification function estimated by probit analysis. Changing the second-formant frequency has not produced significant changes in the duration of silence required to hear a stop. The nonsignificant trend for stops to be heard at shorter durations of silence as the second formant frequency increased may be related to the rise in discriminability of the duration of silent intervals as the spectral contiguity of their bounding markers is increased (for example, Divenyi and Danner, 1977).

The "phonetic" hypothesis can be assessed by comparing the crossover between [s]+vowel and [sp]+vowel responses in Figure 5a with that between [s]+vowel and [sk]+vowel responses in Figure 5d. Although the crossover for the velars occurs at a shorter duration of silence than that for the bilabials, the difference between the positions of the crossovers is not significant. This suggests that there is no simple causative relationship between the phonetic labeling of a stop and the amount of silence required to hear that stop.

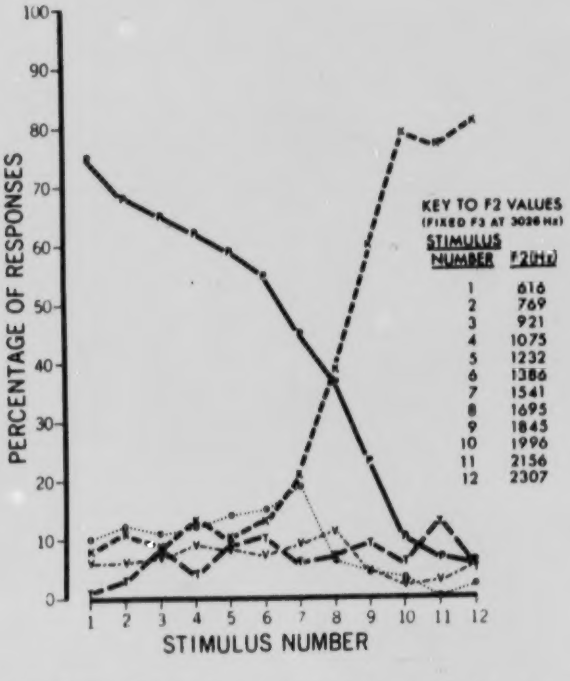
The stimuli used in this experiment specified information for stop place in a closely controlled fashion, and were, therefore, an appropriate starting point for investigating the extreme versions of the acoustic and phonetic hypotheses outlined above. However, no correlation between production and perception emerged, possibly for the very reason that, in these schematic stimuli, the acoustic differences between bilabials and velars were reduced to a minimum. A correlation might emerge with stimuli that reflect more fully the acoustic differences that are naturally concomitant with stops produced at



KEY TO F2 VALUES

STIMULUS NUMBER	F2(Hz)
1	616
2	769
3	921
4	1075
5	1232
6	1386
7	1541
8	1695
9	1845
10	1996
11	2156
12	2307

Figure 1: Identification functions from Experiment I for stimuli incorporating 100 msec of silence. See text for details.



KEY TO F2 VALUES
(FIXED F3 AT 3026 Hz)

STIMULUS NUMBER	F2(Hz)
1	616
2	769
3	921
4	1075
5	1232
6	1386
7	1541
8	1695
9	1845
10	1996
11	2156
12	2307

Figure 2: Identification functions from Experiment II for stimuli incorporating 100 msec of silence. See text for details.

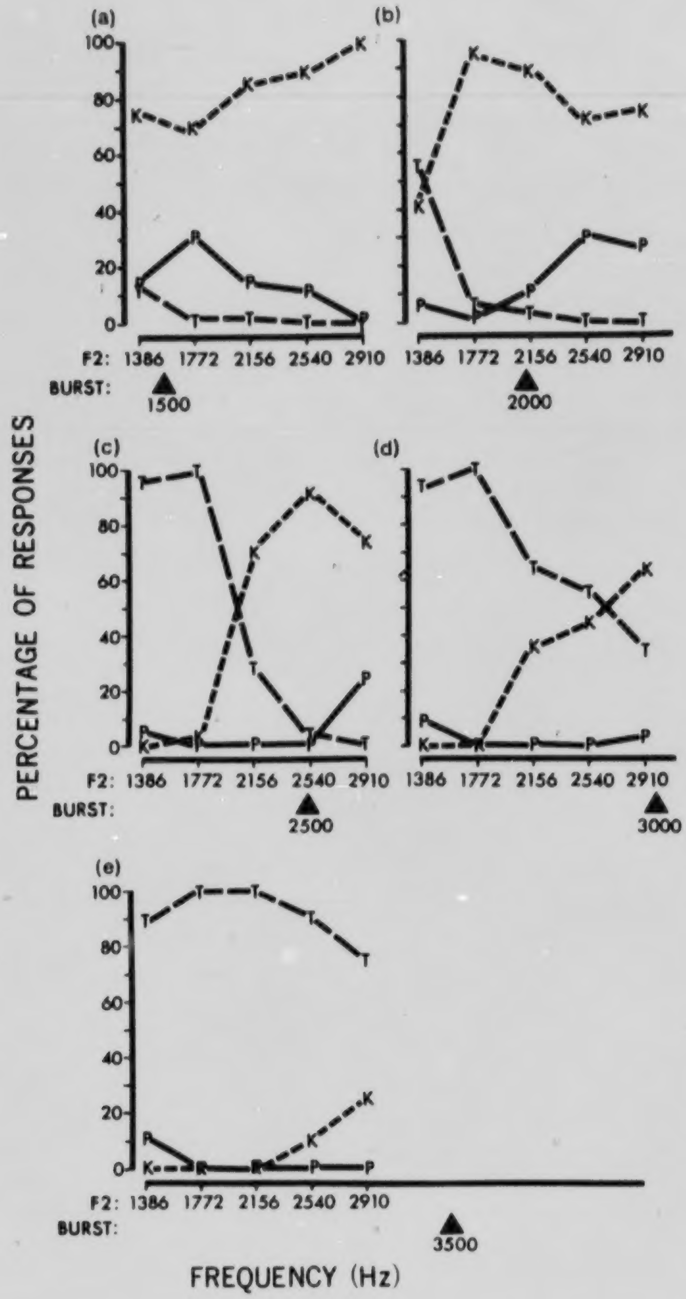
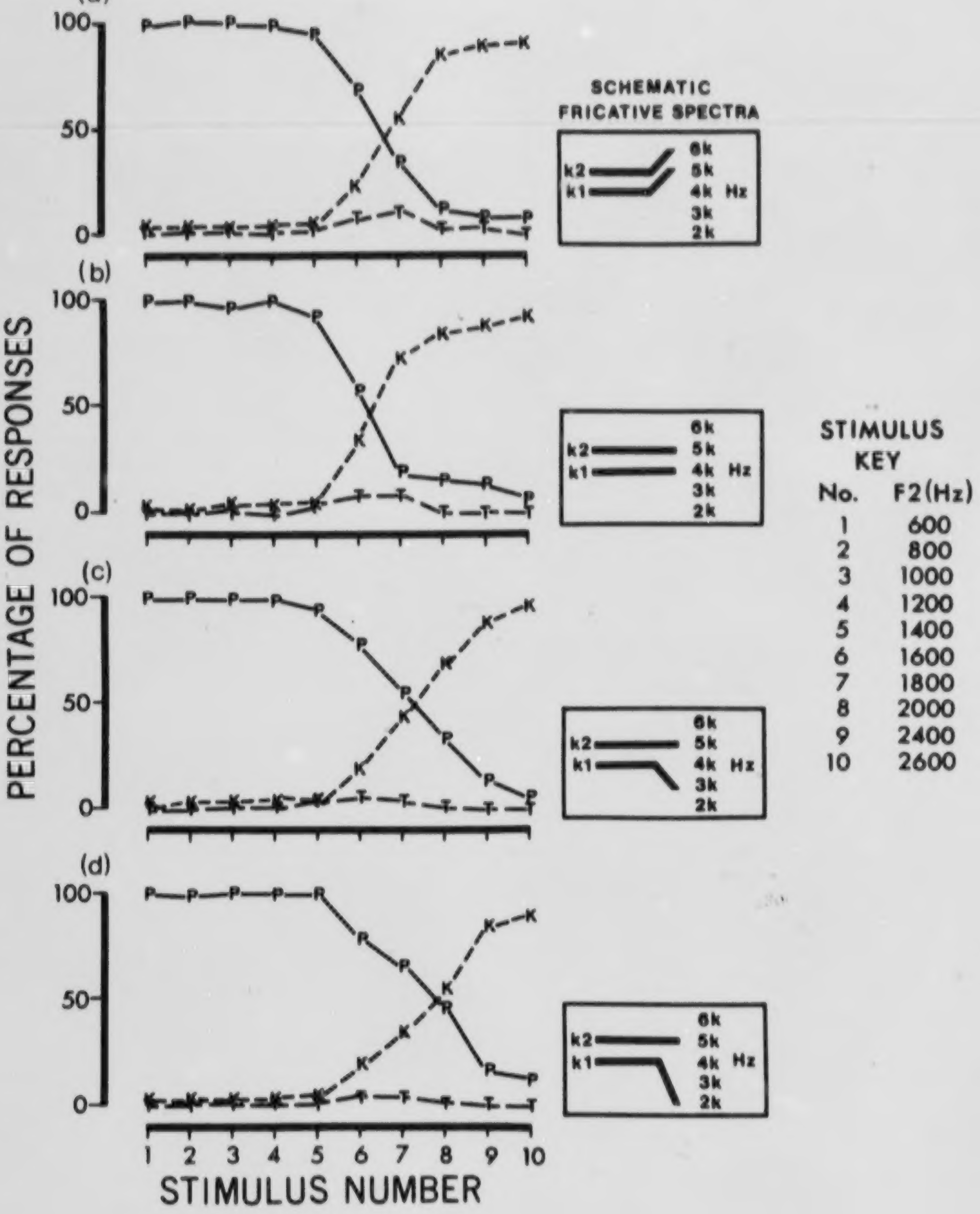


Figure 3: Identification functions from Experiment III. Each panel corresponds to stimuli with the burst frequency indicated by a triangle. The percentages of [p], [t] and [k] responses are plotted against the frequency of the second formant in the vocalic portion of the stimulus.



STIMULUS KEY

No.	F2(Hz)
1	600
2	800
3	1000
4	1200
5	1400
6	1600
7	1800
8	2000
9	2400
10	2600

Figure 4: Identification functions from Experiment IV. See text for details.

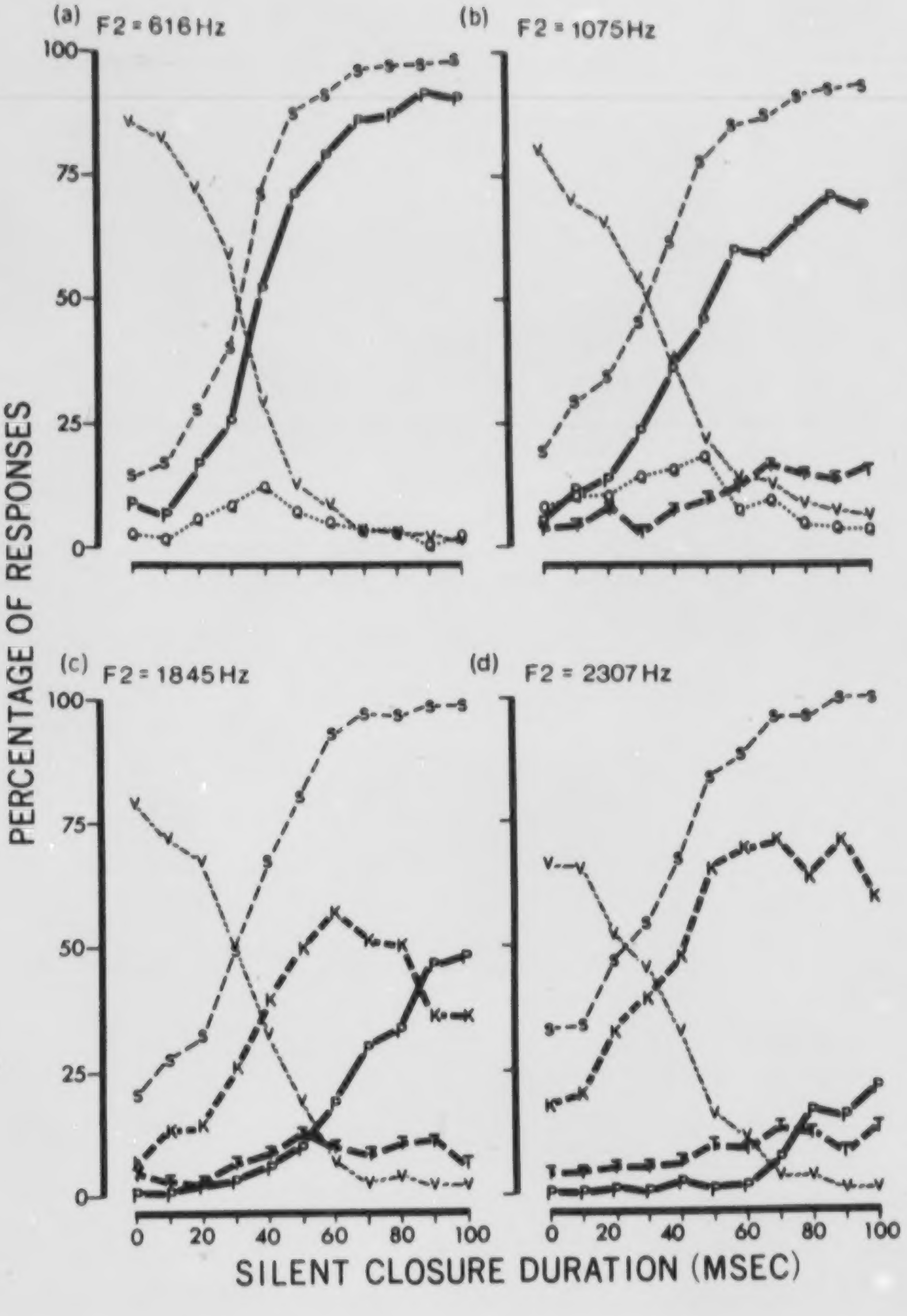


Figure 5: Identification functions from Experiment V. See text for details.

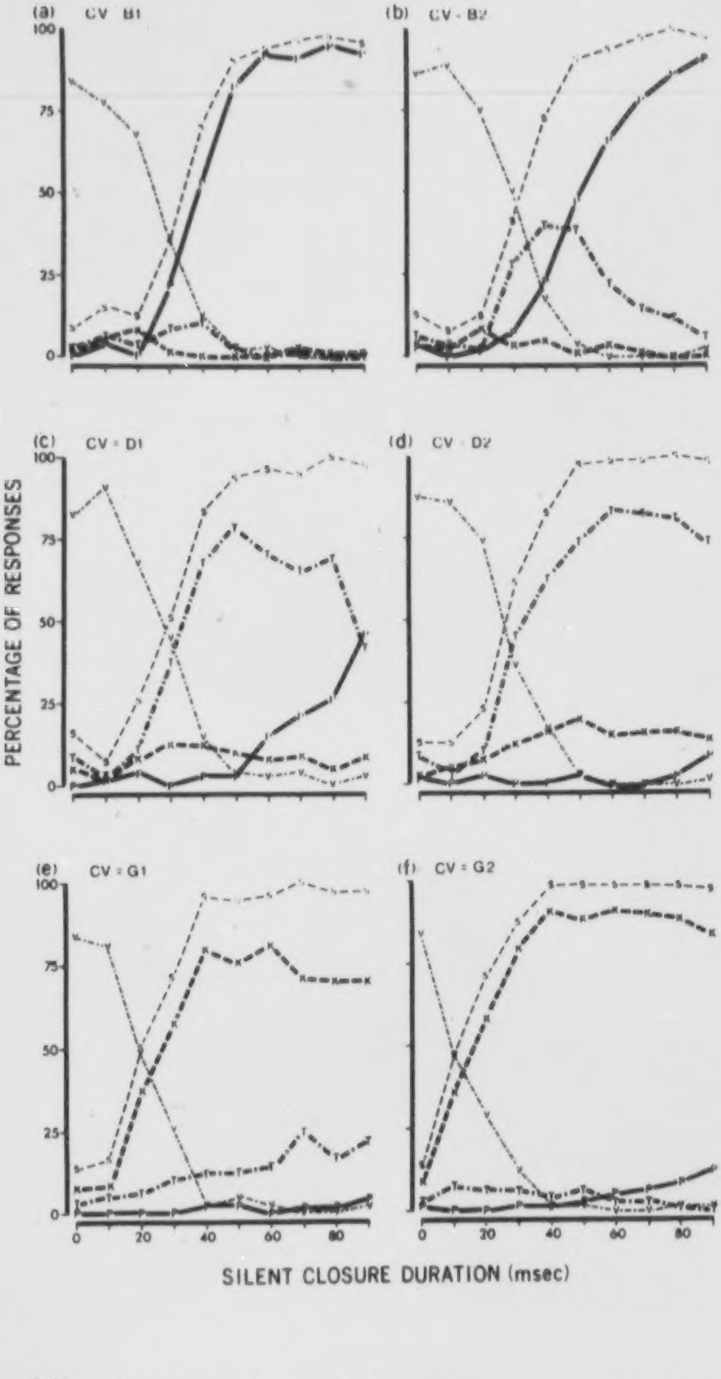


Figure 6: Identification functions for Experiment VI. See text for details.
44

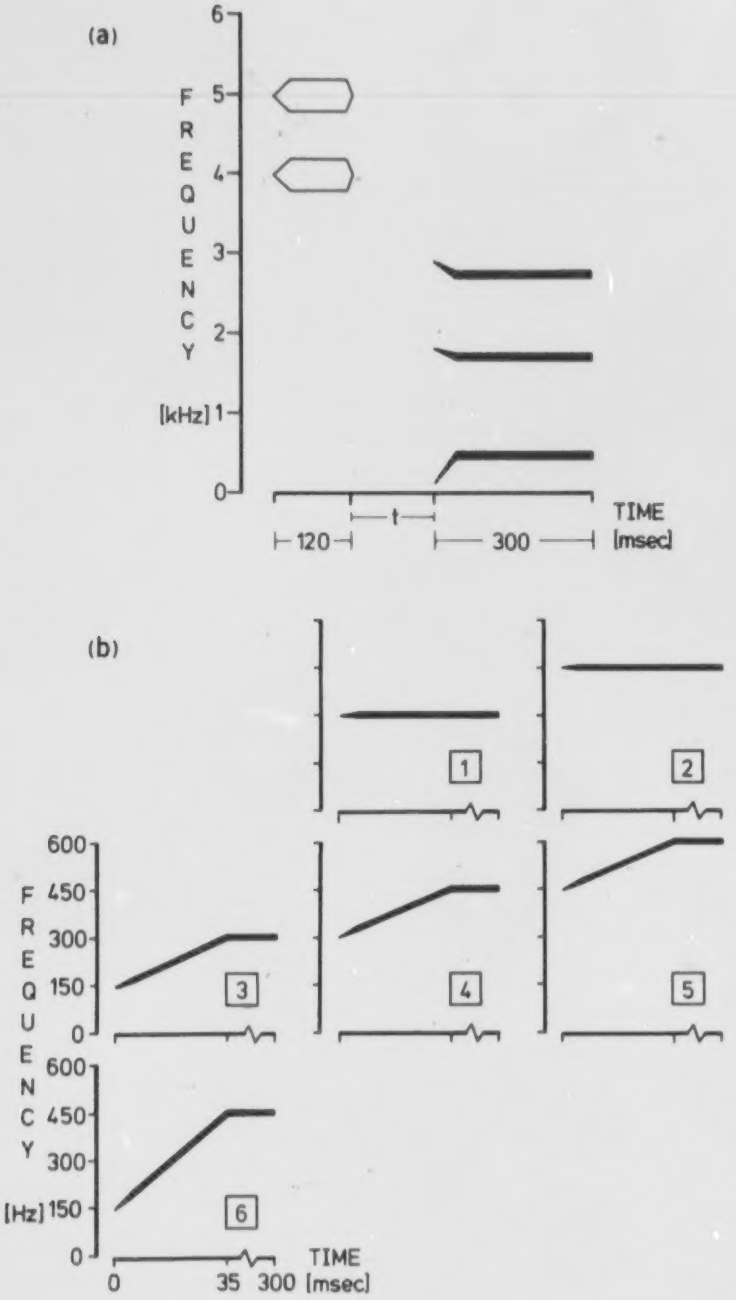

(a)
F R E Q U E N C Y
[kHz]
6
5
4
3
2
1
0
t
120
300
TIME
[msec]
(b)
1
2
3
4
5
6
F R E Q U E N C Y
[Hz]
600
450
300
150
0
TIME
0
35
300
[msec]

Figure 7: (a): Schematic spectrogram of a typical stimulus from Experiment VII. Stimulus series were created by varying the interval t from 0 to 90 msec in 10 msec steps.

(b): Schematic representation of the first formant contours in the stimuli used in Experiment VII. See text for details.

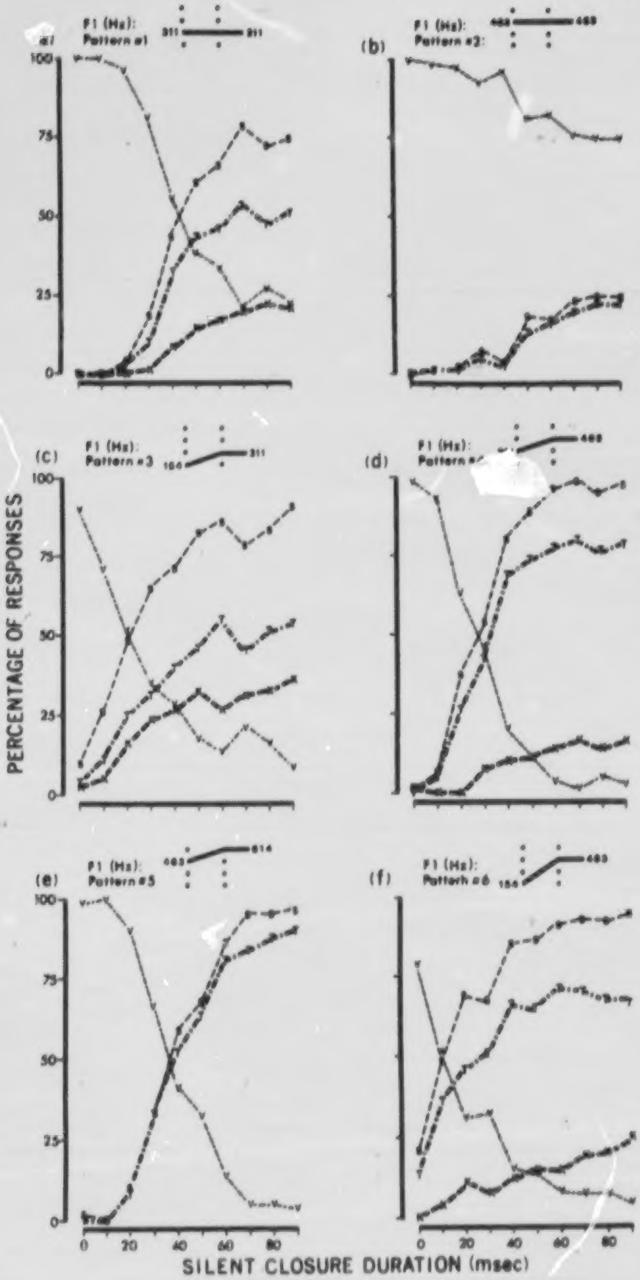


Figure 8: Identification functions from Experiment VII. See text for details.

different places of articulation. Below, in Experiments VI and VII, we investigated the effects of manipulating transitions in the second and third formants, and then transitions in the first formant, on the duration of silence required to hear a stop.

The acoustic and phonetic hypotheses do not exhaust the explanatory principles that can be brought to bear on the results of Experiment V. In making an analytic contrast between two phonetically distinct articulatory events, it is natural to focus first upon the most prominent acoustic consequence of articulation that distinguishes the two events. If manipulation of this single parameter leads to different phonetic percepts, the parameter is accorded the status of a "cue." However, it is commonly found that, if the most potent cues are neutralized by being set to ambiguous values, perceptual sensitivity to less prominent acoustic properties may be demonstrated. Given sufficiently precise control over the acoustic structure of stimuli, it appears to be possible to demonstrate that some "cue-value" attends every acoustic detail that distinguishes two different phonetic events. In the limiting case it becomes clear that every acoustic consequence of an articulatory event is a potential source of information about that event. Thus, according to this rationale, in order to demonstrate perceptual sensitivity to closure duration as a determinant of stop place, we should neutralize all other cues to place of articulation. This situation was approximated in the stimuli for which identification data are displayed in Figure 5c. Their second formant frequency was chosen, on the basis of the results of Experiment I, to give approximately equal numbers of [p] and [k] percepts with a silent interval of 100 msec. Figure 5c shows that both [p] and [k] were perceived. Moreover, other place cues were sufficiently neutral to allow closure duration itself to disambiguate place of articulation: as the production data predict, [k] percepts predominated at short closure durations, while [p] percepts predominated at longer closure durations.

Figure 5c does not show simply that the crossover between [s]+vowel and [sp]+vowel percepts occurred at a longer duration of silence than the crossover between [s]+vowel and [sk]+vowel responses. This result would be a direct consequence of the ambiguous place category being resolved in favour of a consistently greater proportion of velars than bilabials. In such a situation, the function corresponding to the less frequent response would always intersect the [s]+vowel function at a longer silent interval than the function corresponding to the more frequent response. This occurs, for instance, in Figure 5b, where the proportion of [p] to [t] percepts remains approximately constant over the range of closure durations incorporated in the stimuli. Figure 5c shows instead that the ratio of bilabial to velar stops is not fixed, but varies systematically with closure duration, with velars predominating at short closures and bilabials at longer closures. It is unfortunate that the closure durations in the experiment were not extended beyond 100 msec so that the predominance of bilabials at longer closure durations could have been shown more convincingly. However, a more stringent demonstration of the effect can be made. It requires that the identification function corresponding to the less frequently used stop category peak at shorter silent intervals than that corresponding to the more frequently used category. This will be shown in Experiment VI.

In summary, the results of Experiment V suggest that, in the traditional terminology, a silent closure interval is a "cue" both to stop manner and to stop place. Its latter role is revealed when other cues to place of articulation are neutralized.

EXPERIMENT VI

The intention of Experiment VI was similar to that of the previous experiment in its concern with the duration of the silent closure interval required to hear a stop. It was designed to determine how this duration is influenced by the spectral specification of second and third formant transitions introducing the vocalic portion of the stimulus.

Stimuli and Procedure

Six consonant-vowel syllables were prepared with the parallel resonance synthesizer. In each syllable the first formant had its onset at 463 Hz and rose linearly to a steady-state at 614 Hz. The second and third formants had their steady-states at 1845 Hz and 2694 Hz, respectively. The onsets of the second- and third-formant transitions were covaried to produce, in informal listening tests, two instances each of [be], [de] and [ge]. The onset frequencies of the second- and third-formant transitions in these syllables were 1386 Hz and 2525 Hz (B1), 1541 Hz and 2694 Hz (B2), 1695 Hz and 2862 Hz (D1), 1845 Hz and 2862 Hz (D2), 1996 Hz and 2694 Hz (G1), and 2156 Hz and 2525 Hz (G2). The formant transitions were all 40 msec in duration and the total duration of the CV syllables was 300 msec. Six [s]+vowel to [s]+stop+vowel series were created by combining each of these vocalic portions with [s]+silence using the same procedure as in Experiment V. Each series consisted of ten stimuli in which the duration of silence ranged from 0 msec to 90 msec in 10-msec steps. Two identification sequences were recorded. A 24-trial practice sequence included two instances of each of the end-points from the stimulus series. The 300-trial test sequence included five presentations of each of the 60 stimuli. Eight subjects with the same qualifications as those who served in Experiment IV listened first to the practice sequence and then to two presentations of the test sequence. Thus each subject listened to ten presentations of each stimulus. The subjects were instructed to identify the syllables as [s]+vowel or [s]+stop+vowel using the same response categories as in Experiment V.

Results and Discussion

The results of Experiment VI, pooled over subjects, are displayed in Figures 6a - 6f. As in Figure 5, percentages of [s]+vowel (V), [s]+stop+vowel (S), and the breakdown of the stop category into individual functions for [p], [t] and [k], glottal and other stops (Q) are plotted as a function of the duration of the silent interval. Functions are not shown for response categories that received fewer than 10 percent of the total number of responses. The data will be discussed first in terms of the relationship between [s]+vowel and [s]+stop+vowel responses, and then in terms of the particular stop heard.

The closure durations corresponding to the cross-overs between [s]+vowel and [s]+stop+vowel responses in the pooled data of Figure 6 are B1: 30.0 msec, B2: 31.2 msec, D1: 29.0 msec, D2: 27.4 msec, G1:19.8 msec, and G2: 10.1 msec. The significance of any changes in the distributions of [s]+vowel to [s]+stop+vowel responses underlying the differences between these cross-overs was assessed in an analysis of variance that examined the proportions of [s]+vowel responses made by each subject to the ten stimuli in each series combined. Overall, the different stimulus series gave significantly different numbers of [s]+vowel responses ($F_{5,35} = 12.32$; $p < 0.01$). A posteriori comparisons made according to the criteria recommended by Tukey (Winer, 1971) show that series G2 received significantly fewer [s]+vowel responses than any other series, and that none of series B1, B2, D1 or D2 differed significantly from one another. These comparisons confirm the finding of Experiment V, that the duration of silence required to hear a stop is not simply a function of the perceived place of articulation of the stop. Instead, a post-hoc examination of the data shows them to correspond quite closely to an acoustic variable, the frequency separation of the second and third formants at the vocalic onset. As the onset frequencies of F_2 and F_3 approximate one another, the duration of silence required to hear a stop is reduced. The second and third formants in pattern G2 are probably close enough at their onsets to fall within one critical band (Sharf, 1970), and the resulting summation of energy could have specified the vocalic onset of this pattern more distinctively than those of the other five patterns. Energy summation may combine with another class of acoustic effect in which the duration of silence required to hear a stop covaries with the amount of spectral change at the vocalic onset. This hypothesis is explored below in Experiment VII for transitions in the first formant. Here, it would predict equivalent outcomes for patterns B1 and G2. Nonlinear additivity of this effect with that of energy summation could have enhanced the difference between pattern G2 and the other five patterns.

The results of the previous experiment (Experiment V) suggested that the relationship between closure duration and the perceived place of a stop consonant following [s] is most likely to be revealed when other cues to stop place are neutralized. This situation was most closely approximated in the present experiment with series B2 and D1. Overall, series B2 received 66.4 percent [p] responses and 28.7 percent [t] responses, while series D1 received 18.2 percent [p] responses and 69.3 percent [t] responses. The production data in Table 2 show that shorter closure durations characterize alveolar compared to bilabial stops. The patterns of data in Figures 6b and 6c, corresponding to series B2 and D1, respectively, reflect this relationship: [t] percepts predominate at short closure durations, and [p] percepts predominate at longer closure durations. In Figure 6b, the function for [t] reponses intersects the function for [s]+vowel responses at a shorter closure duration than does the function for [p] responses, despite there being a smaller proportion of [t] than [p] responses overall. As was noted in the discussion of Experiment V, this situation provides a convincing demonstration of the relationship between stop closure duration and perceived place of stop articulation. There are two reasons why equivalent effects are not shown in Figures 6d and 6e for series D2 and G1 that straddle the alveolar-velar boundary. First, these patterns specified place of production less ambiguously than did patterns B2 and D1. Of the [s]+stop responses to series D2, 78.7 percent were [t], while 90.3 percent of the [s]+stop responses to series G1 were [k]. Second, in natural productions of [s]+stop clusters, the difference between stop closure durations in alveolars and velars is much smaller than that between either of these categories and bilabials (Table 2).

In summary, the results of Experiment VI confirm those of Experiment V. The duration of the stop closure following [s] can serve to disambiguate bilabial from alveolar and velar place categories when other cues to place of articulation are neutralized. On the other hand, the probability of hearing any stop at a particular closure duration following an [s] is largely determined by the acoustic structure of the vocalic portion of the stimulus. Crossovers betwen [s]+vowel and [s]+stop+vowel responses fell at about the same closure duration on continua whose vocalic portions exemplified [be] and [de]. Crossovers for continua constructed from vocalic [ge] fell at shorter durations. In production, however, alveolars and velars are characterized by similar closure durations, while bilabials are distinguished by longer intervals. If a perceptual trading relationship exists that reciprocates the differences that occur in production across the different places of articulation, it cannot be completely determined by differences in the spectrotemporal specification of F_2 and F_3. Accordingly, the final experiment investigated the influence of the characteristics of the first formant on the duration of silence required to hear a stop after [s]. Possibly, a trading relationship exists between the characteristics of F_1 and the closure duration required to hear a stop. This could compensate both for the differences in closure duration that occur between the different place contexts, and for those, noted earlier in relation to the production data, between more and less open vowels.

EXPERIMENT VII

Experiment VII was designed to determine how the onset frequency and the magnitude of the first formant transition influence the duration of stop closure required to hear a stop after [s].

Stimuli and Procedure

Six CV syllables were created with the parallel resonance synthesizer. They had identical second and third formant contours. The F_2 and F_3 onset at 1541 Hz and 2862 Hz fell linearly to steady-state frequencies of 1312 Hz and 2525 Hz, respectively. All transition durations were 35 msec. The duration of each syllable was 300 msec. A typical stimulus is schematized in Figure 7a. The first formant contours fill six cells of a 3x3 matrix designated by three values of F_1 onset frequency and three values of F_1 transition extent as illustrated in Figure 7b. The transition extent of contours #1 and #2 was 0 Hz, of contours #3, #4 and #5, was 157 Hz, and of contour #6, was 309 Hz. The onset frequency of contours #3 and #6 was 154 Hz, of contours #1 and #4, 311 Hz, and of contours #2 and #5, 463 Hz. Six ten-member [s]+vowel to [s]+stop+vowel stimulus series were constructed, as before, by interpolating silent intervals ranging from 0 msec to 90 msec in 10-msec steps between 120 msec of [s] friction and each vocalic segment. Two identification sequences were recorded. A practice sequence of 24 trials included two instances of the end-point stimuli from each series. A test sequence of 300 trials contained five instances of each of the 60 stimuli.

Eight subjects with the same qualifications as those who had taken part in Experiment IV listened first to the practice sequence and then to two presentations of the test sequence to yield 10 identifications of each stimulus by each subject. The instructions were identical to those given in the two previous experiments.

Results and Discussion

The data from the six stimulus series pooled over the eight subjects are displayed in Figures 8a - 8f. As before, each graph displays the percentages of [s]+vowel responses (V), [s]+stop+vowel responses (S), and the breakdown of the stop category into individual functions for [t] and [k] responses, as a function of the duration of the silent interval. Figure 8 is supplemented by Table 3, in which four summary measures are tabulated for each of the six stimulus series. These are: (a) the duration of silence at the cross-over between [s]+vowel and [s]+stop+vowel responses estimated from the pooled data, (b) the overall percentages of [s]+stop responses, (c) the percentage of [k] responses out of the total number of [s]+stop responses, and (d) the percentage of [t] responses out of the total of [s]+stop responses. Measures (a) and (b) were derived from the data of all eight subjects. Measures (c) and (d) were derived from the data of four subjects who made both [k] and [t] responses.

Considering first the relation between [s]+vowel and [s]+stop+vowel responses, two trends are evident: crossovers between [s]+vowel and [s]+stop+vowel responses occurred at shorter silent intervals both as the onset frequency of F_1 was lowered, and as the magnitude of the F_1 transition was increased (see Table 3a). The statistical significance of these effects could not be assessed directly using the crossover measure, because a single crossover could not be estimated directly for every subject on every stimulus series. As an alternative, directional t-tests were performed on the percentages of [s]+stop responses, a measure that could be determined for every subject. By comparing the pair of percentages in each column of the matrix in Table 3b, an assessment of the effect of increasing the magnitude of the F_1 transition by 154 Hz may be made. Series #6 produced more [s]+stop responses than did series #3 [$t_7 = 2.65$; $p < 0.025$; (1-tailed)]; significant effects in the same direction were found between series #4 and #1 [$t_7 = 5.97$; $p < 0.01$ (1-tailed)], and between series #5 and #6 [$t_7 = 6.60$; $p < 0.01$; (1-tailed)]. Thus, in all three cases, a greater magnitude of F_1 transition produced a larger percentage of [s]+stop responses. Similarly, by comparing pairs of percentages in adjacent rows of Table 3b, there are three comparisons that allow an assessment of the effect of lowering the onset frequency of F_1 by 154 Hz. Series #3 and #4 did not produce significantly different means ($t_7 = 0.15$); however, both series #1 and #2 [$t_7 = 3.90$; $p < 0.01$; (1-tailed)], and series #4 and #5 [$t_7 = 2.30$; $p < 0.05$; (1-tailed)] differed significantly. In each case, a lower F_1 onset frequency produced a larger percentage of [s]+stop responses. Although one of the changes in F_1 onset frequency (from 311 Hz to 154 Hz) failed to produce a significant increase in the percentage of [s]+stop responses, all three were accompanied by a reduction in the duration of silence at the cross-over between [s]+vowel and [s]+stop+vowel responses (Table 3a). Overall, it appears that both of the manipulations applied to the first formant in Experiment VII can produce systematic effects on the duration of silence required to hear a stop after [s].

The response patterns of four of the eight subjects included both [k] and [t] responses, while the other four subjects only made [t] responses. The percentage of [k] and [t] responses out of the total number of [s]+stop responses are displayed in Tables 3c and 3d for the four listeners for whom the place category was ambiguous. Each table shows a consistent trend: the percentage of [sk] percepts increased as the onset frequency of F_1 was lowered and as the magnitude of its transition was reduced, while the inverse pattern applied to [st] percepts. Just as perceptual sensitivity to the covariation of closure duration with place of articulation was demonstrated in Experiments V and VI when other information for place was ambiguous, so here, four subjects have shown perceptual sensitivity to the covariation in production of characteristics of the first formant transition with place. Spectrograms of natural utterances typically show that this covariation involves longer slower first-formant transitions for initial velar stops, that may also entail a lower first-formant onset fequency, than alveolars [for example, Fant (1973), page 118].

Tables 3a and 3b imply the existence of a perceptual trading relationship between the spectral properties of the first-formant transition and the duration of silence required to hear a stop after [s]. Both silence, which is an indicant of a completely constricted vocal tract, and a first-formant rising from a low frequency, which is an indicant of the release of vocal tract constriction, are natural acoustic concomitants of the production of a stop consonant. If it is assumed that a perceptual system for speech exists that is sensitive to both these attributes and that seeks a criterial amount of information for the presence of a stop, then a trading relationship between the two attributes would be expected. Less silence is required when the spectral attribute is more prominent. [See also, Summerfield and Haggard (1977), Erickson, Halwes, Fitch and Liberman (1977)]. The production data in Table 2 endorse the utility of a system organized in this way: the duration of the stop closure is inversely related to the rate at which the oral constriction is released. Thus, shorter closures characterize bilabials compared to alveolars and velars, and, for a given place, shorter closures precede the open vowel [a] compared to the more closed vowels [i] and [u].

DISCUSSION

Summary of Results

The experiments reported here shared a concern for the perception of stop consonants in syllable-initial [s]+stop clusters. The first two experiments showed that the sequence [s]+silence+vowel can give rise to the percept of an initial [s]+stop cluster. The perceived place of the stop was related to the frequency of the second formant: low second-formant frequencies gave [sp] percepts and high second-formants gave [sk] percepts; very few [st] percepts occurred. In contrast, Delattre et al. (1955) reported that in syllable-initial position, where information for stop manner is carried by a rising first formant, a steady second formant at 1800 Hz gives alveolar percepts substantiating the principle of formant loci. In medial position after [s], where information for stop manner is conveyed by a period of silence simulating stop closure, the absence of a significant number of [t] percepts requires consideration of a wider range of the consequences of production than are entailed in the principle of formant loci. The steady-state vocalic portions of the stimuli used in Experiments I and II could only represent natural articulations that give rise either to no release burst, or to contiguous energy at and following the release. The hypothesis was developed that alveolar percepts were absent because the stimuli failed to simulate the spectral discontinuity between alveolar release and the following formant pattern. These notions were tested in Experiment III, where the spectral relationship between a release burst and the following second formant was

Table 3: Results of Experiment VII
Each matrix contains three filled cells relating the Onset Frequency and Transition Extent of the first formant in the vocalic portion of the stimulus series.

Table 3a : Crossovers between [s]+vowel and [s]+stop+vowel responses (msec).

		Onset Frequency (Hz)		
		154	311	463
Transition	0	*	43.6	†90.0
	157	20.8	27.7	36.5
Extent	309	10.6	*	*

Table 3b : Percentages of [s]+stop responses out of all responses (%).

		Onset Frequency		
		154	311	463
Transition	0	*	42.0	7.5
	157	64.6	65.5	54.6
Extent	309	74.9	*	*

Table 3c : Percentages of [sk] responses out of all [s]+stop responses (%).

		Onset Frequency		
		154	311	463
Transition	0	*	41.3	26.7
	157	55.2	24.3	9.5
Extent	309	28.2	*	*

Table 3d : Percentages of [st] responses out of all [s]+stop responses (%).

		Onset Frequency		
		154	311	463
Transition	0	*	48.5	64.4
	157	32.6	65.9	83.6
Extent	309	57.1	*	*

systematically manipulated in [s]+silence+burst+vowel syllables. In accordance with predictions based on the acoustic concomitants of natural productions, the perception of place of articulation in the interpolated stop was determined by the spectral relationship between the burst and second formant: [st] percepts were reported when the burst was at a higher frequency than, and discontinuous with, F_2; [sk] percepts were reported when the burst was spectrally contiguous with F_2; [sp] percepts were infrequent with these concentrated bursts but were sometimes reported when the burst frequency was low. We note that the perceptual data in these experiments are rationalized not by identifying a relationship between perceived place and any particular cue dimension, but by determining the articulatory event whose acoustic consequences are most closely approximated by each stimulus pattern.

The relative success of an appeal to the details of articulation as an explanatory principle for perception motivated Experiments IV to VII. An analysis of natural productions of syllable-initial [s]+stop clusters showed that both the spectral properties of the offset of [s] friction and the duration of the silent closure interval are characteristic of the place of production of the stop. Experiment IV showed perceptual sensitivity to the first of these characteristics: lowering the offset frequency of the fricative predisposes [sp] percepts primarily at the expense of [sk] percepts. Experiments V and VI demonstrated that the duration of the stop closure can determine perceived place of articulation when spectral information for place is ambiguous: shorter closure intervals predispose [st] and [sk] percepts as opposed to [sp] percepts. In a similar fashion, Experiment VII showed that, for some listeners, the characteristics of the first formant at and following the release can determine perceived place of articulation when other information for place is equivocal. Experiment VII also demonstrated an interrelationship between the duration of stop closure and the spectral characteristics of the first formant in the perception of stop manner: a shorter duration of silence is required to hear a stop after [s] as the onset frequency of the first formant is lowered and as the magnitude of its transition is increased. This perceptual trading relationship appears to reciprocate an inverse correlation in production between the duration of stop closure and the openness of the following vowel.

We acknowledge that the stimulus series used in these experiments are not representative of any natural articulatory dimension. Moreover, the schematic vocalic portions of the stimuli were deliberately constrained and, in some cases, involved spectral changes not typical of natural productions. However, the stop percepts in the stimuli were not unnatural, and, when subjects were provided with response categories for ambiguous percepts, they rarely used them. While interpretations of perceptual data obtained with constrained synthetic stimuli must be tempered by these considerations, the consistency of our results with predictions based on analyses of natural productions endorses the utility of the approach for exploring and accounting for the limits of perceptual sensitivity. The technique of using schematic and geometrically specified stimuli is a powerful tool for demonstrating the gross relationship between perceptual identity and articulatory events. However, its failure to represent the subtleties of natural articulatory dynamics limit its ability to generate a complete characterization of the information specifying phonetic identity.

The Concept of a "Cue"

The methodology of the present experiments can be traced directly to early work that sought to specify the acoustic cues of speech (for example, Liberman, Delattre and Cooper, 1952). The techniques of analysis and synthesis provided an operational definition of a cue as a physical parameter of a speech signal whose variation could systematically change the phonetic interpretation of the signal. A large body of data attests to the absence of a one-to-one correspondence between a physical specification of the cues and the phonetic percepts that they induce (for example, Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967). The belief that a more nearly invariant relationship exists between phonetic interpretation and the events of articulation motivated a class of perceptual models that sought to specify how a representation of articulation might be recovered from the substrate of the acoustic cues. Thus, the cue achieved a functional role in a perceptual system as an element of information used in the construction of a featural description of the signal in articulatory terms. This description was assumed to permit a more direct mapping to phonetic identity (for example, Mattingly and Liberman, 1969; Stevens and House, 1972). Even models which did not make the reconstruction of articulation explicit assumed that articulatory reference can sometimes mediate the acoustic-phonetic translation (for example, Pisoni and Sawusch, 1975).

In such models, a particular phonetic feature is detected when a criterial amount of information favoring that feature has been accumulated from the available cues. Trading relationships between cues conveying information for the same feature are inevitable: the greater the amount of information available from one cue, the smaller the amount of information required from another, in order to attain the criterion for deciding that the feature is present. The notion that the information for phonetic identity is conveyed by acoustic cues appeared to possess the attraction of delimiting the critical aspects of the signal, and thereby reducing the amount of information that the perceiver must process. Given the rapid rate at which phones are uttered, it is desirable that the amount of information processing required to perceive each phone be minimized. This was seen to be achieved in part by the parallel transmission of cues as an inevitable result of coarticulation. However, in practice, the adequacy of the account would rest on there being, in addition, a finite, and ideally small, number of cues to be processed for any particular phonetic distinction. The data from the present experiments do not encourage the belief that the set of cues is constrained: for stops after [s], at least, perceptual sensitivity has been demonstrated to each of the acoustic consequences of production that we have chosen to investigate. In the production of a stop consonant, constriction, occlusion and release of the supralaryngeal vocal tract are aspects of a continuous articulatory event that unfolds over time. In a comparison of stops articulated at two different places of production, the configurations of the articulators, and hence the concomitant acoustic signals, are likely to differ between the two events at every moment within their time span. Presumably, given sufficiently precise stimulus control, perceptual sensitivity could be demonstrated to every difference between two articulations; the set of cues to the distinction would then appear to be unbounded. The problem might be resolved by postulating a ranking of the cues in order of importance, such that sufficient information to disambiguate every phone is conveyed by a limited subset of the total number of cues. However, it is a requirement of this solution to the problem that the speech processor ignore the minor cues, at least in the process of speech perception under normal conditions. If this were not the case, part of the perceptual task would be to distinguish major from minor cues, and would require, therefore, that all cues be registered. If it is allowed that the minor cues are normally ignored, it is a further requirement that the major cues completely specify the phonetic contrast that they distinguish, since the same cue can play different roles in different contexts (Liberman et al. 1967). However, if the minor cues are ignored under natural conditions of speech perception, it is paradoxical that perceptual sensitivity can be demonstrated to them at all.

It is appropriate to ask whether a list of any number of cues is the best, or even a sufficient, specification for a phonetic category. That it may not be the best specification is suggested by the fact that different, although not necessarily exclusive, sets of cues can specify the same phone in different contexts. That it cannot be a sufficient specification follows from the fact that, as the experiments reported here have shown, the information for phonetic perception is distributed over time. A set of cues will only be correctly interpreted if they occur with the proper temporal coordination. It follows that, in order to detect a particular phone, the perceptual system must be constrained to register not only a specific set of cues, but also the temporal coordination that articulation imposes upon them. It is noteworthy that while articulatory events occur over time, and therefore that cues are distributed over time, attention has traditionally been directed toward specifying the cues independently of their temporal coordination. As a complement to this approach, perceptual models have typically included an early stage in which the set of cues for a particular phone, coherently arrayed in the speech stream by the events of articulation, are fractionated out as discrete elements, subsequently to be redintegrated by the mediation of internally generated rules of articulatory coherence.

These observations on the cues imply that a perceptual system in which the information for phonetic perception is a set of cues, would have to incorporate three kinds of knowledge if it were to function successfully. It would have to know, first, which aspects of the acoustic signal are cues and which are not; second, it would need to possess a sensitivity to the pattern of co-occurrence of cues for each phone in its perceptual repertoire; third, it would need to appreciate the proper temporal coordination of the cues within each pattern. We can see no reason, in principle, why a device could not be built to perceive phonetic identity from a substrate of acoustic cues, provided it was endowed with an articulatory representation sufficient to embody these three kinds of knowledge. However, we doubt that such a system could evolve in the natural world. For a species to acquire a knowledge of articulatory constraints, it would be necessary first that information specifying those constraints be available for the species, and second that the species possess a prior sensitivity to that information. The knowledge that a particular set of cues combine to indicate the presence of a given phone could be acquired in either of two ways. The identity of the phone could be specified independently of the set of acoustic cues, but this would hardly solve the problem and would preempt the need to evolve a sensitivity to the cues. Alternatively, the signal could directly specify both the identity of the cues and their temporal coordination, but information in the signal that specified the coherence of its elements would, isomorphically, specify the articulatory event from which that coherence derived. However, the presence of this information about articulation in the signal, and a predisposition to register it on the part of the perceiver, would obviate the need for an internalized articulatory referent to mediate the the acoustic-phonetic translation.

These considerations lead us to question the validity of equating the operational and functional definitions of an acoustic cue. A cue was defined operationally as a physical parameter of a speech signal whose manipulation systematically changes the phonetic interpretation of the signal. While it is clear that perceptual sensitivity must exist to the consequences of manipulating a cue, it is not necessary for the cue to be registered in perception as a discrete element.

It follows from the belief that the information for a perceiver is a set of cues that the physical stimulus underdetermines the percept and that, in consequence, the information in the signal must be supplemented by the perceiver's knowledge of the world. In the case of speech perception, this view requires that a knowledge of articulatory constraints mediates the interpretation of the cues. However, the foregoing discussion has argued that this knowledge could only have been acquired if information about articulatory events were directly available in the speech signal. Our inclination is to avoid this paradox by supposing that those aspects of articulation that render speech sounds, as a class distinct from all other sounds, and that serve to distinguish one speech sound from another, are represented in the signal in a veridical, but undoubtedly complex, fashion, and are directly available to the perceiver. This is clearly not a solution to the problem of understanding speech perception, but it argues for a change in orientation to the problem. The task remains that of specifying why the same percept arises from the distinct acoustic patterns produced when the same phone is articulated in different contexts. A critical part of the attainment of this specification is to determine a level of description at which the linguistically relevant events of articulation and the acoustic signal are isomorphic. A characterization of this isomorphism will define the information in speech for a perceiver and should facilitate a solution to the traditional alternatives to theories of perception that posit the construction of percepts from a substrate of cues can be found elsewhere (for example, Gibson, 1966; Turvey, 1977).

The general viability of this orientation is supported by two types of experimental observation. The first is found in recent demonstrations that place of articulation is more directly represented in the acoustic signal than has been supposed hitherto (Kuhn, 1975; Stevens and Blumstein, 1977). The second type of observation is that there exists perceptual sensitivity to the higher order properties of events that is not dependent on an initial fractionation of the stimulus into discrete elements. For instance, it has been shown that velocity may be apprehended directly in vision without the prior mediation of representations of displacement and time (Lappin, Bell, Harm and Kottas, 1976). This demonstration of direct perceptual sensitivity to change over time suggests that the perception of events in general, including articulatory events, may involve the direct apprehension of change over time, and may, therefore, not require the perceptual integration of a succession of discrete cues. Consistent with this suggestion is the fact that vowels uttered in dynamic (CVC) context are perceived more accurately than are tokens of the same vowels produced in isolation. The result implies that a complete specification of the information for the perception of a vowel may entail a specification of change over time[4] (Strange, Verbrugge, Shankweiler and Edman, 1976; Shankweiler, Strange and Verbrugge, 1977).

The data from the experiments reported here do not confirm, but certainly encourage, the belief that the perception of speech is the perception of information isomorphic with articulatory events and not the assimilation of a succession of discrete cues. This kind of experimentation would be more fruitful if it were the complement to studies of the relationship between articulation and acoustics, so that stimulus patterns could be specified not in the arbitrary metric of Euclidean geometry, but in the natural metric of articulatory dynamics. As a result, the endeavour might demonstrate perceptual sensitivity to the coherence in the acoustic consequences of articulatory events, that is, to the information underlying the experience of phonetic perception.

REFERENCES

Bastian, J. (1959) Silent intervals as closure cues in the perception of stop phonemes. Paper presented to 58th Meeting of the Acoustical Society of America, Cleveland, Ohio.

Cooper, F. S., P. C. Delattre, A. M. Liberman, J. M. Borst, and L. J. Gerstman. (1952) Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America 24*, 597-606.

Delattre, P. C., A. M. Liberman, and F. S. Cooper. (1955) Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America 27*, 769-773.

Divenyi, P. L. and W. F. Danner. (1977) Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception & Psychophysics 21*, 125-142.

Dorman, M. F., M. Studdert-Kennedy, and L. J. Raphael. (1977) Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics 22*, 109-122.

Erickson, D. M., H. L. Fitch, T. G. Halwes, and A. M. Liberman. (1977) Trading relation in perception between silence and spectrum. *Journal of the Acoustical Society of America 61*, S46(A).

Fant, G. (1960) *Acoustic Theory of Speech Production*. (The Hague: Mouton).

Fant, G. (1973) *Speech Sounds and Features*. (Cambridge, Mass: M. I. T. Press).

Gibson, J. J. (1966) *The Senses Considered as Perceptual Systems*. (Boston: Houghton-Mifflin).

Klatt, D. H. (1974) The duration of [s] in English words. *Journal of Speech and Hearing Research 17*, 51-63.

Klatt, D. H. (1975) Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research 18*, 686-706.

Kuhn, G. M. (1975) On the front cavity resonance and its possible role in speech perception. *Journal of the Acoustical Society of America 58*, 428-

---

[4] Fowler, C. A. (1972) Timing control in speech production. Doctoral dissertation, University of Connecticut (unpublished).

433.

Lappin, D. S., H. H. Bell, O. J. Harm, and B. Kottas. (1975) On the relation between time and space in the visual discrimination of velocity. Journal of Experimental Psychology: Human Perception and Performance 1, 383-394.

Liberman, A. M., F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. (1967) Perception of the speech code. Psychological Review 74, 431-461.

Liberman, A. M., P. C. Delattre, and F. S. Cooper. (1952) The role of selected stimulus-variables in the perception of the unvoiced stop consonants. American Journal of Psychology 65, 497-516.

Lotz, J., A. S. Abramson, L. J. Gerstman, F. Ingemann, and W. J. Nemser. (1960) The perception of English stops by speakers of English, Spanish, Hungarian and Thai: A tape-cutting experiment. Language and Speech 3, 71-77.

Mattingly, I. G. and A. M. Liberman. (1969) The speech code and the physiology of language. In Information Processing in the Nervous System, ed. by K. N. Leibovic. (New York: Springer Verlag), 97-117.

Peterson, G. E. and H. L. Barney. (1952) Control methods used in a study of the vowels. Journal of the Acoustical Society of America 24, 175-184.

Pisoni, D. B. and J. R. Sawusch. (1975) Some stages of processing in speech perception. In Structure and Process in Speech Perception, ed. by A. Cohen and S. Nooteboom. (Heidelberg: Springer-Verlag), 16-34.

Raphael, L. J., M. F. Dorman, A. M. Liberman. (1976) Some ecological constraints on the perception of stops and affricates. Journal of the Acoustical Society of America 59, S25(A).

Reeds, J. A. and W. S-Y. Wang. (1961) The perception of stops after [s]. Phonetica 6, 78-81.

Shankweiler, D. P., W. Strange, and R. R. Verbrugge. (1977) Speech and the problem of perceptual constancy. In Perceiving, Acting and Knowing: Towards an Ecological Psychology, ed. by R. Shaw and J. Bransford. (Hillsdale, N. J.: Lawrence Erlbaum Associates), 315-346.

Sharf, B. (1970) Critical bands. In Foundations of Modern Auditory Theory, ed. by J. V. Tobias, vol. 1. (New York: Academic Press), 157-202.

Stevens, K. N. (1975) The potential role of property detectors in the perception of consonants. In Auditory Analysis and the Perception of Speech, ed. by G. Fant and M. A. A. Tatham. (New York: Academic Press), 303-330.

Stevens, K. N. and S. E. Blumstein. (1977) Onset spectra as cues for consonantal place of articulation. Journal of the Acoustical Society of America 61, S48(A).

Stevens, K. N. and A. S. House. (1956) Studies of formant transitions using a vocal tract analog. Journal of the Acoustical Society of America 28, 578-585.

Stevens, K. N. and A. S. House. (1972) Speech perception. In Foundations of Modern Auditory Theory, ed. by J. V. Tobias, vol. 2. (New York: Academic Press), 3-62.

Strange, W., R. R. Verbrugge, D. P. Shankweiler and T. R. Edman. (1976) Consonant environment specifies vowel identity. Journal of the Acoustical Society of America 60, 213-221.

Summerfield, A. Q. and M. P. Haggard. (1977) On the dissociation of spectral and temporal cues to the voicing distinction in initial stops. Journal of the Acoustical Society of America 62, 435-453.

Turvey, M. T. (1977) Contrasting orientations to the theory of visual information processing. Psychological Review 84, 67-88.

Weimer, W. B. (1977) A conceptual framework for cognitive psychology: Motor theories of the mind. In *Perceiving, Acting and Knowing: Towards an Ecological Psychology*. ed. by R. Shaw and J. Bransford. (Hillsdale, N.J.: Lawrence Erlbaum Associates), 267-314.

Winer, B. J. (1971) *Statistical Principles in Experimental Design.* (New York: McGraw-Hill).

Perceptual Integration of Acoustic Cues for Stop, Fricative, and Affricate Manner

Bruno H. Repp, Alvin M. Liberman[†], Thomas Eccardt[††], and David Pesetsky[†††]

ABSTRACT

As is well known, introducing a short interval of silence between the words SAY and SHOP causes the listener to hear SAY CHOP. Another cue for the fricative-affricate distinction is the duration of the fricative noise in SHOP (CHOP). Now, varying both these temporal cues orthogonally in a sentence context, we find that, within limits, they are perceived in relation to each other: the shorter the duration of the noise, the shorter the silence necessary to convert the fricative into an affricate. On the other hand, when the rate of articulation of the sentence frame is increased while holding noise duration constant, a longer silent interval is needed to hear an affricate, as if the noise duration, but not the silence duration, were effectively longer in the faster sentence. In a second experiment, varying noise and silence durations in GRAY SHIP, we find that, given sufficient silence, listeners report GRAY CHIP when the noise is short, but GREAT SHIP when it is long. Thus, the long noise in the second syllable disposes the listener to displace the stop to the first syllable, so that he hears not a syllable-initial affricate (that is, stop-initiated fricative), but a syllable-final stop (followed by a syllable-initial fricative). Repeating the experiment with GREAT SHIP as the original utterance, we obtain the same pattern of results together with only a moderate increase in GREAT responses. In all such cases, the listener integrates a numerous, diverse, and temporally distributed set of acoustic cues into a unitary phonetic percept. These several cues have in common only that they are the products of a unitary articulatory act. In effect, then, it is the articulatory act that is perceived.

[†] Also University of Connecticut and Yale University.

[††] Also Yale University.

[†††] Also Massachusetts Institute of Technology, Cambridge.

Acknowledgment: This research was supported by NICHD Grant HD01994. We would like to thank Georgann Witte for her help in data analysis. A short version of this paper was presented at the 94th Meeting of the Acoustical Society of America in Miami Beach, December 1977.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 2 (1978)]

INTRODUCTION

When a speaker makes an articulatory gesture appropriate for a phonetic segment, the acoustic consequences are typically numerous, diverse and distributed over a relatively long span of the signal. In the articulation of an intervocalic stop consonant, for example, the characteristically rapid closing and opening of the vocal tract has acoustic consequences that include, among others, the following: various rising and falling transitions of the several formants; a period of significantly reduced sound intensity; and a second, acoustically different set of formant transitions plus, (in the case of voiceless stops in iambic stress patterns) a transient burst of sound, a delayed onset of the first formant, and, for the duration of that delay, band-limited noise in place of periodic sound in the higher formants. These many acoustic features are somehow integrated into the perception of a single stop consonant, though they are, as is plain, extremely diverse in character and distributed, sometimes, over periods as long as 300 msec (Repp, 1976).

To account for the integration of these cues, we find it reasonable to suppose that they are processed by a system specialized to perceive the phonetically significant act by which they were produced. On that assumption, we should expect that all of the cues associated with such an act would result in a unitary percept, as indeed they do. The boundaries of the integration would then be set, not by the number, diversity, or temporal distribution of the cues, but rather by a decision that they do (or do not) plausibly specify an articulatory act appropriate for the production of a single phonetic segment.

In the experiments to be reported here, our aim is to learn more about the ways in which acoustic cues produce integral phonetic percepts. To that end, we have, in the first experiment, examined the integration of two temporal cues--duration of silence and duration of fricative noise--in the perception of the distinction between fricative and affricate; and we have also investigated the effect on that integration of a still more widely distributed temporal variable, namely, the rate at which the surrounding speech is articulated. In the second experiment, we have studied the effects of those same temporal cues, but now in connection with the perception of juncture. That provides us with an opportunity to examine a case in which the integration occurs across syllable boundaries: a syllable-final stop is perceived or not, depending on a cue in the next syllable that simultaneously determines whether the initial segment in that syllable is taken to be a fricative or an affricate.

EXPERIMENT I

In this experiment, we selected two cues for study, both temporal in nature and both relevant to the fricative-affricate distinction. One of them is silence. A short period of silence (or near-silence) in the acoustic signal tells the listener that the speaker has closed his vocal tract, a gesture characteristic of stop consonants and affricates. That silence is a powerful and often sufficient cue for the perception of stop or affricate manner can be experimentally demonstrated by inserting silence at the appropriate place in an utterance. So, for example, SLIT can be converted into a convincing SPLIT by inserting a sufficient amount of silence between the fricative noise and the vocalic (LIT) portion. That was done originally in tape-splicing experiments (Bastian, 1959, 1960; Bastian, Eimas and Liberman, 1961). For the same phonetic contrast, investigators have more recently explored the range of effective silence durations (Dorman, Raphael, and Liberman, 1976) and, in another study, revealed a trading relation between silence and a spectral cue.[1] Other contrasts--similar in that they, too, are based on the presence or absence of stop or stop-like manner--have also been found to depend in important ways on the silence cue. Thus, with appropriate insertions of silence, SI can be made to sound like SKI or SU like SPU (Bailey and Summerfield, 1978). Silence can also be sufficient to cue the fricative-affricate contrast in, for example, SAY SHOP vs. SAY CHOP (Dorman et al., 1976); it is this contrast that will concern us here.[2]

For the fricative-affricate contrast, there are, as always, other cues besides silence. The one we have used in our experiment is duration of (fricative) noise, a cue shown originally by Gerstman (1957) to be important. Thus, we have two temporal cues, duration of silence and duration of noise. To those two temporal cues, we have added a variable that is also temporal in nature: rate of articulation. Our interest in observing the effects of that variable springs from several sources. We might expect, first of all, that the effect of articulatory rate would be especially apparent on cues that are themselves durational in nature. Several studies tend to confirm that expectation (for example, Pickett and Decker, 1960; Ainsworth, 1974; Fujisaki, Nakamura, and Imoto, 1975; see also Footnote 3). Indeed, one of these studies deals with the same fricative-affricate contrast we mean to study and reports a seemingly paradoxical effect: having determined that increasing the duration of silence between SAY and SHOP was sufficient to convert the utterance PLEASE SAY SHOP to PLEASE SAY CHOP, Dorman et al. (1976) found that when the rate of the precursor PLEASE SAY was increased, more silence was needed to produce the affricate in CHOP. We wish to test for that effect at each of several durations of the fricative noise, and in a larger sentence context.

Our motivation to make those tests stems from the possible bearing of the results on an interpretation of the way a listener adjusts to changes in the acoustic stimulus patterns caused by variations in rate of speech. That interpretation may be complicated in interesting ways if, as has been reported by students of speech production (for example, Kozhevnikov and Chistovich, 1965), changes in rate of articulation do not stretch or compress all portions of the speech signal proportionately. In that connection, the data most relevant to our purposes are owing to Gay (1978). He found that durations of silence associated with stop consonants change less with rate than do the durations of the surrounding vocalic portions. It is possible, then, that the cues of our experiment--duration of silence and duration of fricative noise--are differentially affected by changes in speaking rate, though we are not aware of any direct evidence for this. At all events, we think it appropriate to investigate further such differential effects that may appear in perception.

---

[1] Erickson, D., H. L. Fitch, T. G. Halwes and A. M. Liberman. (1978) A trading relation in perception between silence and spectrum. Unpublished manuscript.

[2] It may be noted that the stop consonants (affricates) in the three examples given have different places of articulation. Perceptual information about place of articulation is provided by spectral cues preceding and following the silence (Bailey and Summerfield, 1978). In our experiments, we are concerned only with cues for stop manner and not with place distinctions. Therefore, we will pass over the question of why, in the last example, listeners hear SAY CHOP (SAY TSHOP) and not SAY PSHOP or SAY KSHOP.

[3] Summerfield, A. Q. (1978) On articulatory rate and perceptual constancy in phonetic perception. Unpublished manuscript.

Method

Subjects. Seven paid volunteers (Yale undergraduates) participated, as well as three of the authors (BHR, TE, DP). All except BHR were native speakers of American English. (BHR learned German as his first language.) The results of all ten subjects were combined, since there were no substantial differences among them.

Stimuli. A male talker recorded the sentence, WHY DON'T WE SAY SHOP AGAIN, at two different speaking rates, using a monotone voice and avoiding emphatic stress on any syllable. The fast sentence lasted 1.26 sec, while the slow sentence lasted 2.36 sec--a ratio of 0.53 . The sentences were low-pass filtered at 4.9 kHz and digitized at a sampling rate of 10 kHz. This was done with the Haskins Laboratories Pulse Code Modulation (PCM) system. Monitoring the waveforms on high-resolution oscillograms, we excerpted the SH-noise of the slow utterance (110 msec in duration) and substituted it for the SH-noise in the fast utterance (originally 92 msec). Thus, the two utterances had identical noise portions.

Knowing that rate of onset of the fricative noise is an important cue for the fricative-affricate distinction (Gerstman, 1957; Cutting and Rosner, 1974), we were concerned that it be neutralized. Preliminary observations suggested that the noise onset in our stimuli was, in fact, not neutral, but rather so gradual as to bias the perception strongly toward fricative and even, perhaps, to override the effects of the two duration cues we wished to study. To remove, or at least reduce, that bias, we removed the initial 30 msec of the noise, leaving 80 msec. That maneuver had the effect of creating a more abrupt onset.[4]

---

[4] This manipulation merely created a situation favorable for obtaining the desired effect and in no way affected the validity of the experiment. In fact, our cutting back the noise resulted in a moderate bias in the opposite direction--toward hearing affricates (CHOP). It should be noted in this connection that not only does SAY SHOP turn into SAY CHOP when silence is inserted, but a natural SAY CHOP can also be turned into SAY SHOP by removing the silence that precedes the fricative noise. Both effects have limits, however: a noise with an extremely abrupt onset will not easily be heard as a fricative even in the absence of silence, and a noise with an extremely gradual onset will not easily be heard as an affricate, even if sufficient silence is present.

We used the PCM system to vary the two temporal cues under study, noise duration and silence duration. Three different noise durations were created by either duplicating or removing 20 msec from the center of the 80-msec noise, leaving its onset and offset unchanged. Thus, the noise durations were 60, 80, and 100 msec in both sentence frames. In each of the resulting six sentences, varying amounts of silence were inserted before the fricative noise. Silence duration was varied from 0 to 100 msec in 10-msec steps. Eleven silence durations, three noise durations, and two speaking rates resulted in 66 sentences. These were recorded in five different randomizations, with 2 sec intervening between successive sentences.

In order to determine how the different noise durations were perceived outside the sentence context, we prepared a separate tape containing isolated SHOP (CHOP) words excerpted from the test sentences. (The stimuli consisted of the portion from the beginning of the fricative noise to the beginning of the P-closure.) Three different noise durations and two speaking rates yielded six stimuli; these were duplicated ten times and recorded in a random sequence, separated by 3-sec intervals. The different speaking rates were reflected in the durations of the vocalic portions of the test words; they were 140 msec (slow) and 113 msec (fast).

Procedure

The subjects listened in a quiet room over an Ampex Model 620 amplifier-speaker, as the tapes were played back on an Ampex AG-500 tape recorder. Intensity was set at a comfortable level. All subjects listened to the isolated words first, except for the three authors; they took this brief test at a later date. The task was to identify each word as either SHOP or CHOP, using the letters S and C for convenience in writing down the responses and guessing when uncertain. The same responses were required in the sentence test. The listeners were informed about the different speaking rates but not about the variations in noise and silence duration (except for the authors). After a pause, the sentence test was repeated, so that 10 responses per subject were obtained for each sentence.

Results

First consider the results obtained for isolated words. Although the original utterance had contained SHOP, the isolated words were predominantly perceived as CHOP. Presumably, this was a consequence of our having cut back the original fricative noise, thus creating not only a shorter noise duration but also a more abrupt onset; both changes would be expected to bias perception towards affricate manner (Gerstman, 1957). Despite the bias, there was a clear effect of the variations in noise duration: The percentages of CHOP responses to the three noise durations (60, 80, and 100 msec) were 99, 91, and 81 (slow rate) and 99, 90, and 73 (fast rate), respectively. Thus, as expected, the probability of hearing an affricate decreased as noise duration increased. In addition, there seemed to be a slight effect of vowel duration at the longest noise duration, again in the expected direction: When the vocalic portion was shorter--this being the only manifestation of the faster speaking rate in the isolated words--the probability of hearing CHOP was lower, indicating that the noise duration was, to some extent at least, effectively longer at the fast speaking rate.

We turn now to the results of the main experiment. That silence was an effective cue for the fricative-affricate distinction in sentence context is shown in Figure 1. There we see that the listeners heard SHOP or CHOP depending on the duration of the silence that separated the fricative noise from the syllable (SAY) immediately preceding it. This replicates earlier findings (Dorman et al., 1976). If, as is reasonable, we consider an affricate to be a stop-initiated fricative, then our result is also perfectly consistent with those of other investigators who have found silence to be important in the perception of stop-consonant manner.

We see, further, that duration of fricative noise had a systematic effect, as indicated by the horizontal displacement of the three functions in each panel of Figure 1. The proportion of SHOP responses increased significantly with noise duration ($F_{2,18} = 32.36$, $p \ll .01$). That effect establishes a trading relationship between silence and noise duration: as noise duration increases, more silence is needed to convert SHOP into CHOP.[5]

The effect of speaking rate can be seen by comparing the two panels of Figure 1. We see that the paradoxical effect first discovered by Dorman et al. (1976) was indeed replicated: for equivalent noise durations, more silence was needed in the fast sentence frame than in the slow sentence frame to convert the fricative into an affricate ($F_{1,9} = 16.35$, $p < .01$).

The foregoing results are represented more concisely in Figure 2. The data points shown there are the SHOP-CHOP boundaries (that is, the 50 percent crossover points of the six labeling functions) as estimated by the method of probits (Finney, 1971). This procedure fits cumulative normal distribution functions to the data; it also yields estimates of standard deviations and standard errors of the boundaries.[6] To show the trading relationship between the temporal cues more clearly, Figure 2 plots the SHOP-CHOP boundaries (abscissa) as a function of noise duration (ordinate) and speaking rate (the two separate functions). Each function describes a trading relationship between noise duration and silence duration by connecting all those combinations of silence and noise durations for which SHOP and CHOP responses are equally probable. The joint dependence of perceptual judgments on both durational cues is indicated by the fact that the trading functions are neither perfectly vertical nor perfectly horizontal, but have intermediate slopes. Both functions are strikingly linear.

---

[5] Strictly speaking, the term "trading relationship" may not be appropriate for a positive relationship between two cues, but we will use the term for want of a better one. The positive covariation of the two perceptual cues is a direct consequence of their negative covariation in production: fricatives have a long noise duration and no silence, while affricates have a shorter noise duration preceded by a closure interval. Genuine perceptual trading relationships (negative covariation) are observed when two acoustic properties are positively correlated in production, such as, for example, silence and the extent of the first-formant transition as cues for stop manner (Bailey and Summerfield, 1978). In any case, a positive trading relationship can be turned into a negative one by simply reversing the directionality of the scale on which one of the cues is measured.

[6] The boundary estimates obtained from the average data of all subjects were virtually identical with the averages of the estimates for individual subjects, so the former have been plotted in Figure 1. The response function for the longest noise seemed to asymptote below 100 percent CHOP responses, especially at the fast speaking rate. This caused the estimated boundaries to fall at somewhat longer silence durations than the 50 percent intercepts shown in Figure 1.

While an increase in speaking rate left the linear form of the trading relationship unchanged, it shifted the function toward longer silence durations, simultaneously changing its slope. This indicates that rate of articulation had a differential effect on the effective silence duration and on the effective noise duration. In fact, the trading functions in Figure 2 coincide well with straight lines through the origin of the coordinate system, which means that, within each speaking rate condition, the fricative-affricate boundary is associated with a constant ratio between silence and noise duration--approximately 0.44 at the slow rate and 0.55 at the fast rate. A separate analysis of variance of silence/noise ratios showed only a significant effect of speaking rate ($F_{2,18} = 14.60$, $p < .01$); the effect of noise duration and the interaction term were far from significant. Thus, the consequence of changing the rate of articulation was a change in the ratio of silence to noise required for the same phonetic perception.[7]

Discussion

It is not novel to find that variations in rate of articulation have an effect on the perception of temporal cues in speech. Nor is it entirely novel to find, as we have, that variations in rate have an unequal effect on the several temporal cues--duration of silence and duration of noise--that are effective in the perception of the fricative-affricate distinction; as we pointed out in the Introduction, that conclusion was suggested by an experiment done by Dorman et al. (1976). We have extended that finding. Having varied both the duration of silence and the duration of noise, we saw that the inequality is not peculiar to a particular duration of noise, and we saw, moreover, a trading relation between the two duration cues. That trading relation now becomes a component of one interpretation of the seemingly paradoxical rate effect.

To appreciate that interpretation in its broadest form, we should note once again the comments by several students of speech production that variations in rate of articulation do not affect all portions of the speech signal equally. To the extent that this is so, a listener cannot adjust for rate variations by applying a simple scale factor, but rather must make a more complex correction--one that embodies a tacit knowledge, as it were, of the inequalities in the signal that rate variations generate. Perhaps the results of our experiment are an instance of that correction and that tacit knowledge. Suppose that, in the case of utterances like those of our experiment, variations in rate of articulation cause the duration of the fricative noise to change more than the duration of the silence. If the listener's perception reflects an accurate understanding of that inequality, then he should expect that, given an increase in rate, the noise would shorten more than the silence. However, on hearing, as in some of the conditions of our experiment, that the noise duration remains constant when the rate increases, the listener would assign to the noise an effectively greater (relative) length. As we know, a longer noise duration biases the perception towards fricative, though, as shown by the trading relation in our results, that bias can be overcome by an increase in the duration of silence. A consequence of all that would be just the effect of rate we found in our experiment: when the rate was increased as the duration of noise was held constant, listeners required more silence to perceive an affricate.

---

[7] It must be kept in mind that this description is true only within the limits of the present experiment. Had the noise duration been increased beyond 100 msec, a point would have been reached where no amount of silence would have led to a substantial percentage of CHOP responses (cf. Experiment II).

The foregoing interpretation depends, among other considerations, on a determination that variations in rate do, in fact, produce the particular inequality that concerns us here. As we pointed out earlier, Gay (1978) found in utterance types somewhat analogous to ours, that rate variations produced smaller variations in the silence associated with stop consonants than in the durations of the surrounding vocalic portions. Unfortunately, there are no data on exactly those utterances we used in our experiment. We have made efforts in that direction, but the results so far are inconclusive. Until such time as we know more clearly just what happens in speech production, the interpretation we have offered here is, of course, quite tentative.

The interpretation must be tentative for yet another reason: it does not reckon with the possibility that certain other cues for the fricative-affricate distinction might have been at work in ways that we do not yet thoroughly understand. We have in mind, particularly, the rise-time of the fricative noise. From the work of Gerstman (1957) and Cutting and Rosner (1974), we know that it is a relevant cue. We do not know, however, exactly how it trades with the two duration cues. More important, we do not know how, or even whether, it varies with rate of articulation. Information on these matters will surely affect our interpretation.

It is of interest to wonder how the rate of articulation was specified by the context in which the fricative-affricate segments were embedded. Did the listener take a kind of running average over long sections of the utterance, or did he, alternatively, rely on rate cues in the immediate environment of the target segments? There is nothing in our experiment that enables us to answer that question. We should, however, take note of a relevant finding, together with an interesting discussion of the matter, by Summerfield.[8] Having discovered that perception of the voice onset time

---

[8] See Footnote 3.

(VOT) cue to the voicing contrast (for stop consonants in initial position) was significantly affected by the rate of articulation, Summerfield was able to determine that the effect was quite local: almost all of the effect could be accounted for by variations in the durations of the target syllable and the syllable immediately preceding it.

In that connection, we should again consider the original finding by Dorman et al. (1976) of the differential effect of speaking rate on the two duration cues for the fricative-affricate contrast. The point is that the effect they measured (for the one noise duration they used) was similar in magnitude to ours, though the target word SHOP was held constant and only the two-syllable precursor PLEASE SAY was presented at fast and slow rates. Thus, the vocalic portion preceding the silence may have been the primary mediator of the speaking-rate effect. Examination of the stimuli of the present experiment revealed a substantial durational difference in the vocalic portions of SAY (180 vs. 100 msec at the slow and fast rates, respectively), so that there was a clear acoustic basis for a "local" speaking rate effect. From a psychoacoustic viewpoint, our finding that the noise cue was more affected by speaking rate than the silence cue is the more surprising, since the vocalic portion preceding the silent interval (in SAY) varied much more with speaking rate than the vocalic portion following the fricative noise (in SHOP)--their respective changes in duration being 80 and 27 msec.

There are two important results of our experiment. One has to do with the trading relationship between duration of silence and duration of noise as joint cues for the fricative-affricate distinction. It is provocative that these cues, diverse and distributed as they seem, are nevertheless integrated into the unitary phonetic percept we call fricative or affricate. In our view, this integration occurs because cues such as these converge through a single decision process that takes account of their common origin: they are the consequences of the same articulatory act. The other result, that we have already discussed at some length, is that the two duration cues were affected unequally by a change in rate of articulation. We would now simply emphasize the inequality, which is a very reliable effect, for it does imply, details of interpretation aside, that perceptual correction for variations in rate is not made in this case by applying a simple scale factor, but may rather require some more sophisticated computation.

EXPERIMENT II

While exploring the boundaries of the phenomenon reported in Experiment I, we observed an effect that we have undertaken to investigate more systematically in Experiment II. We reported in Experiment I that, with increases in the duration of silence between SAY and SHOP, the fricative in SHOP changed to the affricate in CHOP. However, when the fricative noise was at its longest (100 msec), it occasionally seemed that CHOP changed back to SHOP, while the stop-like effect was displaced to the end of the preceding syllable, converting SAY to SAYT. If confirmed, that effect would be interesting because it bespeaks an integration of perceptual cues across syllable (word) boundaries. It is also relevant to the problem of "juncture", so long a concern of linguists. (See Lehiste, 1960.)

Acoustic data about juncture, obtained by analysis of the speech signal, have been available for some time, but experimental manipulations of the candidate cues have only recently been undertaken. Two of the experimental investigations are pertinent to the one we will report here. In one of these studies (Christie, 1974), it was shown that placement of a syllable boundary in the string ASTA was affected both by the duration of the silence associated with the stop consonant and also by whether or not the stop was aspirated. More relevant, perhaps, is a study by Nakatani and Dukes (1977). They investigated the role of various cues by cross-splicing portions of natural utterances contrasting in the position of juncture (for example, PLAY TAUGHT vs. PLATE OUGHT), and they concluded that "what we hear at the end of a word ... depends on how the next word begins" (p. 718). In other words, the cues in the initial portion of the second word determined where listeners located the word boundary. For our experiment, these results imply that the duration of the fricative noise at the beginning of the second syllable is likely to be a major cue for the perceived location of juncture. In fact, Lehiste (1960) reported that the fricative noise in natural utterances of WHITE SHOES was considerably longer than that in WHY CHOOSE--a contrast similar to that employed in our present experiment.

Our concern, then, is with the cues that affect perception and placement of stop-consonant manner, either as a final segment added to one syllable or as the conversion of the first segment of the next syllable from fricative to affricate. The cues we have examined are the same as those of Experiment I, duration of silence between the syllables and duration of the fricative noise at the beginning of the second syllable, but with two changes. In order to offer maximum opportunity for the stop-like effect to be transferred from the second syllable to the first, we have included durations of fricative noise longer than those used in Experiment I, thus providing a stronger bias against affricate percepts. To make the alternative responses equally plausible to our subjects, we used a new sentence, DID ANYBODY SEE THE GRAY (GREAT) SHIP (CHIP). The sentence context was employed to make the test as natural as possible. (Rate of articulation was not a variable in this experiment.)

In a second part of the experiment (Experiment IIb), we assessed the effects of those spectral and durational cues that distinguish GRAY and GREAT. For that purpose, we investigated how the results depend on whether, in the original recording, the word was pronounced as GRAY or as GREAT.

Method

Subjects. The subjects were the same as in Experiment I.

Stimuli: Experiment IIa. The sentence, DID ANYBODY SEE THE GRAY SHIP, was produced by a male speaker in a monotone voice and recorded in digitized form. Using the editing facilities of the Haskins Laboratories PCM System, we varied the duration of silence inserted before the word SHIP from 0 to 100 msec in steps of 10 msec. The duration of the fricative noise in SHIP was also varied. Starting with the duration of the noise as recorded, which was 122 msec, we excised or duplicated 20-msec portions from its center, thus shortening or lengthening it without changing the characteristics of its

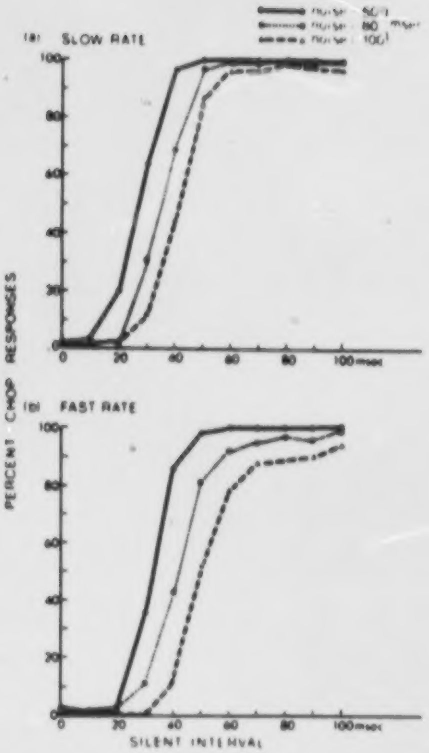


Figure 1: The effect of duration of silence and duration of fricative noise on the perceived distinction between fricative (SHOP) and affricate (CHOP). This is shown for each of the two rates at which the sentence frame was articulated.

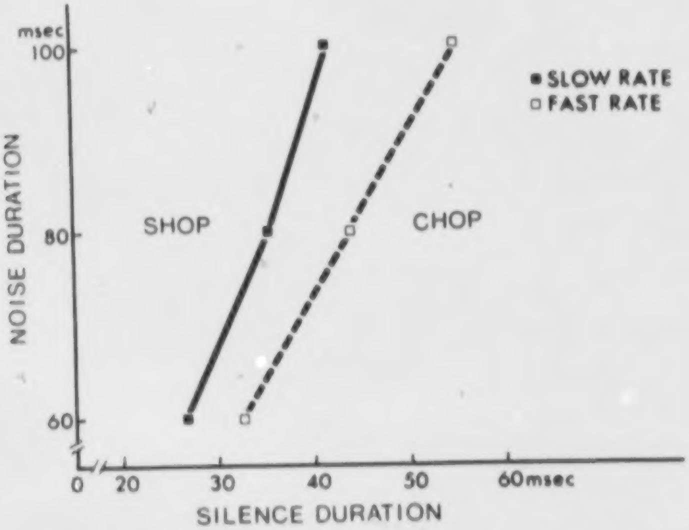


Figure 2: Boundaries between perceived fricative (SHOP) and affricate (CHOP) at each speaking rate as joint functions of the duration of silence and the duration of fricative noise.

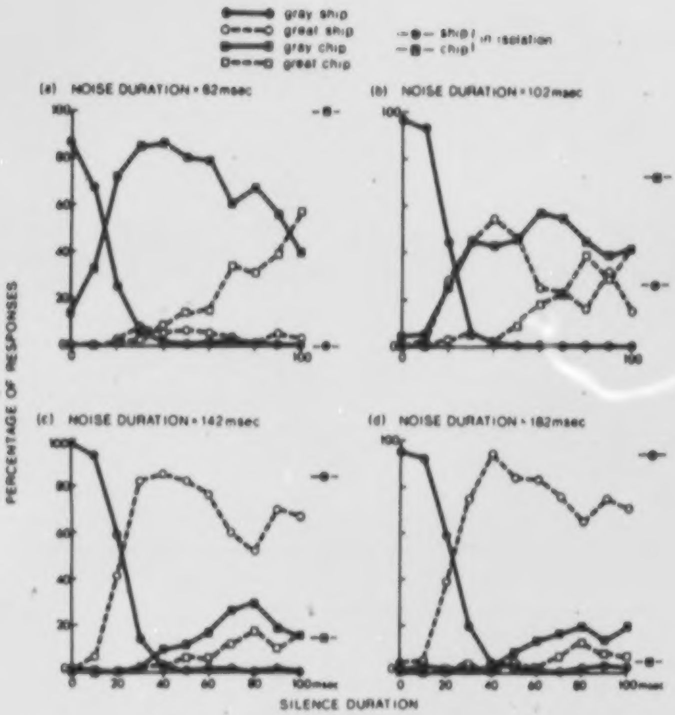


Figure 3: The effect of duration of silence, at each of four durations of fricative noise, on the perception and placement of stop (or affricate) manner.

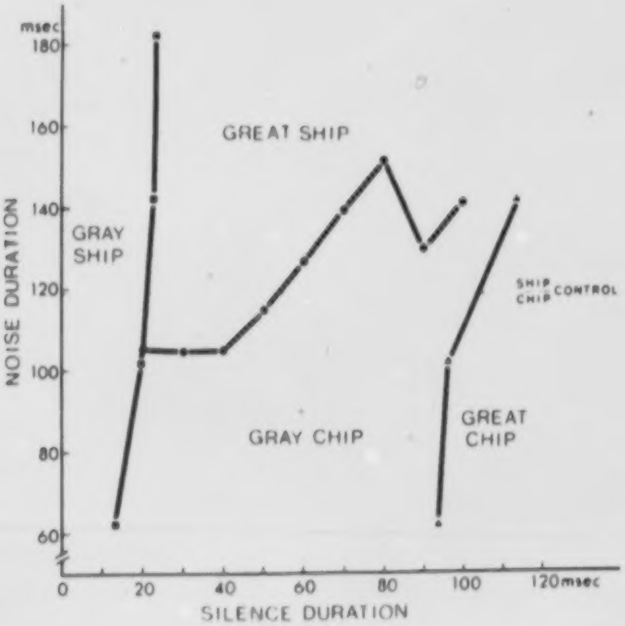


Figure 4: Boundaries that divide the several response categories, represented as joint functions of duration of silence and duration of fricative noise.

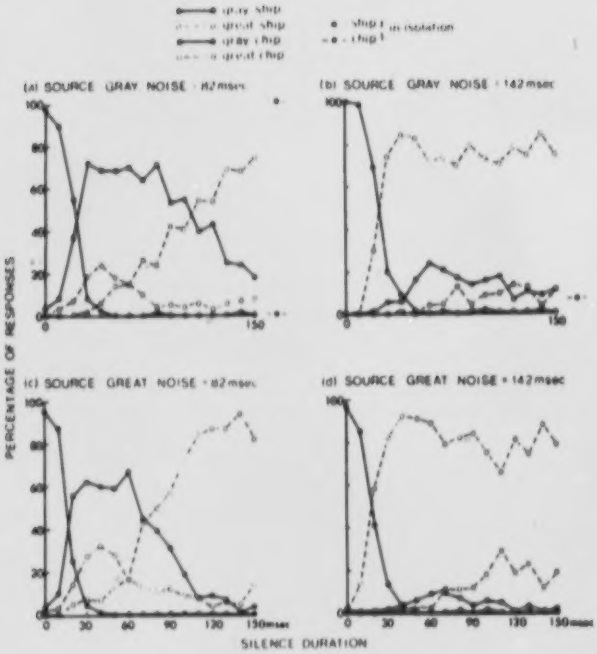


Figure 5: The effects of varying the "source" (original pronunciation as GRAY or GREAT) on the perception and placement of stop (or affricate) manner. These are shown at each of two durations of noise, and represented as the percentages of occurrence of the several responses plotted against the duration of silence.

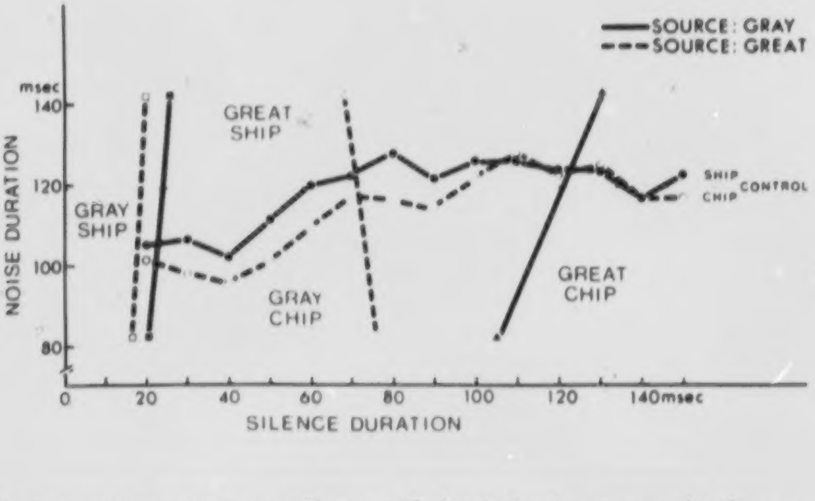


Figure 6: The effects of varying the "source" (original pronunciation as GRAY or GREAT) on the boundaries that divide the several response categories.

onset or offset. In this way, we created four durations of noise--62, 102, 142, and 182 msec--for use in the experiment. Four noise durations and eleven silence durations led to 44 test utterances. These were recorded in five different randomizations, with intervals of 2 sec between sentences.

In order to see how the fricative-affricate distinction is affected by noise duration alone, we excised the word SHIP (CHIP) and varied the duration of the noise as described above, but in steps of 20 rather than 40 msec. These isolated words were recorded in a randomized sequence containing 10 repetitions of each stimulus. The interstimulus interval was 3 sec.

Stimuli: Experiment IIb. A second sentence, DID ANYBODY SEE THE GREAT SHIP, was recorded by the same speaker who had produced the sentence, DID ANYBODY SEE THE GRAY SHIP, of Experiment IIa. He attempted to imitate the intonation and speaking rate of the first-produced sentence. That he succeeded well was suggested by our own listening and by comparison of the waveforms. Using the PCM System, we excerpted the fricative noise from the SHIP of Experiment IIa and substituted it for the noise in the corresponding word of the new sentence. Thus, the two stimulus sentences had exactly the same fricative noise in the final word SHIP. Both sentences were used in Experiment IIb: the original sentence, DID ANYBODY SEE THE GRAY SHIP, and the new sentence, DID ANYBODY SEE THE GREAT SHIP; the important difference was simply in the opposition between the words GRAY and GREAT.

Inspection of waveforms and spectrograms revealed that there was only a slight difference in duration between the two utterances; this difference was almost entirely accounted for by the additional closure period between GREAT and SHIP in the second sentence. The transitions of the second and third formants were, as expected, somewhat steeper in GREAT than in GRAY. Also, the GREAT syllable had a longer duration (210 msec, not including the following closure period) than GRAY (187 msec).[9] Their offset characteristics were similar.

Only two noise durations, 82 and 142 msec, were used, as against the four (62, 102, 142, and 182 msec) of Experiment IIa. There were more silence durations, on the other hand, covering the (wider) range from 0 to 150 msec in steps of 10 msec. Thus, with two noise durations, 16 silence durations, and two sentence frames, there were 64 test sentences in all. These were recorded in five randomized sequences.

---

[9] Our intuition may tell us that GRAY should have been longer than GREAT. However, this intuition is based on the pronunciation of these words in isolation, where word-final lengthening extends the vowel in GRAY. When followed by SHIP, on the other hand, the longer duration of GREAT is quite plausible. However, we do not know whether this observation has any generality.

Procedure

Experiments IIa and IIb were conducted in a single session of about 2 hours duration. The isolated word sequence was presented first (the response alternatives being SHIP and CHIP), followed by the sentences of Experiments IIa and IIb, in that order. Each set of sentences was repeated once, so that each subject gave 10 responses to each sentence. The subjects chose from four response alternatives, using letter codes in writing down their responses: A = GRAY SHIP, B = GREAT SHIP, C = GRAY CHIP, D = GREAT CHIP. No subject had any difficulties using this system.

Results

Experiment IIa: Figure 3 shows the effects of the two cues, duration of silence and duration of fricative noise, on the perception of stop or stoplike manner in the utterance DID ANYBODY SEE THE GRAY (GREAT) SHIP (CHIP). Duration of silence is the independent variable; the four panels correspond to the durations of fricative noise. At the right of each panel, we have also shown the results obtained when the second of the key words, SHIP (CHIP), was presented in isolation.

Let us consider first the responses to the isolated word SHIP (CHIP). At noise durations of 62, 102, 142, and 182 msec--those used in the experiment--the percentages of CHIP responses were 100, 73, 16, and 6, respectively. Thus, as we had every reason to expect, duration of the noise is a powerful cue for the fricative-affricate distinction. The SHIP-CHIP boundary was estimated to be at 119 msec of noise duration. In contrast to the stimuli of Experiment I, whose noise durations all fell below this boundary and therefore were predominantly heard as affricates, those of the present experiment spanned the entire range from affricate to fricative.

The more important results of the experiment are seen by examining the graphs that tell us how the stimuli were perceived in the sentence context. We note first that, when the silence was of short duration--less than 20 msec--the subjects perceived primarily GRAY SHIP. At those very short durations of silence no stoplike effect was evident, either as an affricate at the beginning of the second syllable (CHIP) or as a stop consonant at the end of the first syllable (GREAT). With increasing durations of silence, a stoplike effect emerged. As in Experiment I, somewhat more silence was required at the longer noise durations for this stoplike effect to occur ($F_{3,27} = 6.93$, $p < .01$).

Perhaps the most interesting result was that, once a stop was heard, its perceptual placement in the utterance depended crucially on the duration of the fricative noise: at short noise durations, the listeners reported predominantly GRAY CHIP; at longer noise durations, GREAT SHIP. This resulted in a significant response category by noise duration interaction ($F_{9,81} = 71.52$, $p << .01$).

We also see that the response percentages were in fair agreement with the results for isolated words: when the critical word was heard as CHIP in isolation, it was generally also heard as (GRAY or GREAT) CHIP in sentence context--provided, of course, that it was preceded by at least 30 msec of silence--while words heard as SHIP in isolation were generally heard as

(GREAT) SHIP. Responses in the GREAT CHIP category occurred at the longer silence durations when the noise was short, but even at the longest silence duration and shortest noise, such responses reached only about 50 percent.

A more concise representation of the results, showing perceptual boundaries as determined by the probit method, is to be found in Figure 4. There we see three functions, each of which links those combinations of silence duration and noise duration that are precisely balanced between certain response alternatives, as we will specify below. The dashed horizontal line represents the SHIP-CHIP boundary for isolated words.

Consider first the nearly vertical function at the left (squares). This function characterizes the boundary between GRAY SHIP and all other responses. In other words, at each combination of silence and noise duration on this function, listeners were just as likely to hear a stoplike effect as they were to hear no stop at all. The lower part of this function, which represents the boundary between GRAY SHIP and GRAY CHIP, corresponds directly to the SAY SHOP--SAY CHOP boundary functions of Experiment I (see Figure 2). As in Experiment I, this part of the function is slanted and thus reflects a trading relationship between silence and noise duration. Moreover, again in agreement with Experiment I, the trading relationship can be described as a constant ratio of silence to noise. However, this ratio--about .20--is considerably smaller than that obtained in Experiment I at a comparable speaking rate (.44). This is presumably due to the fact that, in the present experiment, less silence was needed to obtain a stoplike effect. The reason why less silence was necessary was suggested by listening to the words preceding the silence when taken out of context. The SAY of Experiment I actually sounded like SAY (not SAYT) in isolation, but the excised word GRAY of the present experiment, although correctly pronounced in the original sentence, sounded much more like GREAT. Thus, the vocalic portion preceding the silence contained stronger stop-manner cues in the present experiment than in Experiment I, so that less silence was required to hear a stoplike effect. These observations provide indirect evidence for yet another trading relationship between two cues for stop manner: the (spectral and temporal) characteristics of the vocalic portion preceding the silence, and silence duration itself.

Returning to the boundary function at the left of Figure 4, we note that the function changes from slanted at short noise durations to completely vertical at longer noise durations. In other words, the trading relationship between silence and noise duration that characterizes the GRAY SHIP vs. GRAY CHIP distinction disappears as the distinction changes to GRAY SHIP vs. GREAT SHIP. This phonetic contrast, located in the first syllable, is apparently not affected by further increases in noise duration in the second syllable, but depends only on silence duration.

We turn now to the second function in Figure 4--that connecting the circles. This function represents the boundaries between GREAT SHIP on the one hand, and GRAY CHIP and GREAT CHIP on the other hand. (GRAY SHIP responses did not enter into the calculation of these boundaries.) Since GREAT CHIP responses occurred primarily at long silence durations, the major part of the boundary function represents the distinction between GREAT SHIP and GRAY CHIP, that is, the perceived location of juncture. Clearly, noise duration was the major juncture cue, as we should have expected given the observations of Lehiste (1960) and Nakatani and Dukes (1977). Had it been the only cue, the boundary function would have been perfectly horizontal. As we see, however, the function shows a clear rise at intermediate silence durations (40-80 msec): GREAT SHIP responses were more frequent at short silence durations, while GRAY CHIP responses were more frequent at longer silence durations. Thus, silence duration was a secondary cue for the location of the word boundary (see Christie, 1974, for a related result).

The third function in Figure 4--that connecting the triangles--represents the boundary between GRAY CHIP and GREAT CHIP, excluding other responses. There was no obvious dependency of this boundary on noise duration; the uppermost data point, which may suggest that such a dependency was based on only a few observations, since, at this noise duration (142 msec), GREAT SHIP responses predominated (see Figure 3). We note that a fairly long period of silence (about 100 msec) was required to hear both a syllable-final stop and an affricate.

Experiment IIb: By using the sentence containing the word GRAY as the "source" for half of the stimuli, Experiment IIb partially replicated Experiment IIa. These results are shown in the top panels of Figure 5. They may be contrasted with the results obtained with the new GREAT source shown in the bottom panels. For each source, the effects of noise and silence duration were similar to those observed in Experiment IIa; they therefore need no further comment. The change in the response pattern as a function of noise duration was again highly significant ($F_{3,27} = 58.95$, $p << .01$).

The effect of primary interest was that of source. It can be seen that more GREAT (both GREAT SHIP and GREAT CHIP) responses occurred when the source was GREAT, as shown by a significant interaction between source and response categories ($F_{3,27} = 10.11$, $p < .01$). However, this effect did not substantially change the overall response pattern. At silence durations of less than 20 msec, the listeners still reported GRAY SHIP; and at longer silence durations GRAY CHIP was heard when the noise was short, even though the original utterance had been GREAT. Thus, the cues for stop manner in the word GREAT were readily integrated with the initial consonant of the next word if the short noise biased perception toward hearing an affricate.

As in Experiment IIa, we have calculated three kinds of perceptual boundaries (see Figure 4).[10] These are shown in Figure 6, where they are plotted, separately for each "source", as joint functions of silence duration and noise duration. We see that the boundary between GRAY SHIP and the other responses (squares) shifted significantly to the left as the source changed from GRAY to GREAT ($F_{1,9} = 33.66$, $p < .01$). In other words, less silence was needed to hear a stoplike effect (regardless of whether it was placed at the end of the first or at the beginning of the second syllable) when the original utterance had contained the word GREAT.[11] Note that the stop manner cues preceding a relatively short silence were readily integrated with those following the silence: within the range of silence (and noise) durations where the subjects' responses were either GRAY SHIP or GRAY CHIP, the frequency of GRAY CHIP responses actually was increased when the source was changed from GRAY to GREAT.

---

[10] The GREAT SHIP vs. GRAY CHIP (+ GREAT CHIP) boundary estimates were based on only two data points (noise durations). In order to obtain probit estimates, two hypothetical anchor points were added: 22 msec (of noise) with 0 percent GREAT SHIP responses, and 202 msec (of noise) with 100 percent GREAT SHIP responses.

The second boundary function--that separating GREAT SHIP from GRAY CHIP and GREAT CHIP responses (circles)--also showed an interesting pattern of source effects. At shorter silence durations, where the distinction was mainly between GREAT SHIP and GRAY CHIP, the change in source from GRAY to GREAT increased GREAT SHIP responses and decreased GRAY CHIP responses. This is reasonable, although it provides a counterexample to the recent conclusion by Nakatani and Dukes (1977) that cues in the first word have no effect on the perceived location of the word boundary. At long silence durations (beyond 100 msec), on the other hand, the phonetic distinction was primarily between GREAT SHIP and GREAT CHIP, and there source ceased to have any effect. Thus, when the silent interval exceeded about 100 msec, stop-manner cues preceding the silence were no longer integrated with those that followed it, about 100 msec.

The third boundary. GRAY CHIP vs. GREAT CHIP (triangles), showed by far the largest source effect. Since the phonetic contrast was located here in the word that was actually changed in pronunciation, and since, because of the relatively long silence duration, the stop manner cues preceding the silence were perceived independently of the cues following it, the large effect is readily understandable. On the other hand, the effect is not trivial, since, as we pointed out earlier, the word GRAY from the GRAY source actually sounded like GREAT in isolation. That the stimuli derived from the GRAY source received any GREAT CHIP responses at all was probably due to the presence of relatively strong stop manner cues in the word GRAY.

---

[11] This effect is in agreement both with the speaking rate effect in Experiment I and with the difference in silence/noise ratios between Experiment I and Experiment IIa. To see why, we recall that there were both spectral and durational differences between the words GRAY and GREAT, as pronounced in the source utterances. GREAT was longer in duration than GRAY--a difference that, in another context, might have resulted from a slower speaking rate; and, as we have seen in Experiment I, less silence is required to hear a stop when the speaking rate is slow than when it is fast. Thus, the interpretation, proposed earlier, that the speaking rate effect in Experiment I was "locally" mediated by the duration (and perhaps other characteristics) of the vocalic portion preceding the silence is in agreement with the source effect found in the present study, to the extent that the latter was due to the durational differences in the vocalic portion. Second, we have noted earlier that less silence was required in Experiment IIa than in Experiment I for a stop or affricate to be heard, and the presumed reason for this was the fact that the word GRAY (Experiment IIa) sounded like GREAT in isolation, while SAY (Experiment I) sounded like SAY. Thus, the vocalic portion preceding the silence conveyed stronger stop manner cues in Experiment IIa than in Experiment I, leading to a corresponding reduction in the silence required. Consequently, a further increase in the strength of stop manner cues, brought about by actually having the speaker pronounce GREAT (rather than GRAY) in the original utterance, should have further reduced the amount of silence required to hear a stoplike effect, as it did. Whether the spectral or the durational differences between GRAY and GREAT were the primary mediators of the source effect cannot be determined from the present data, but this is an interesting question for further research.

Discussion

The most interesting aspect of the data, in our view, is that whether or not a syllable-final stop consonant was perceived (GRAY vs. GREAT) depended on the duration of the noise following the silence--an acoustic event occurring much later in time. There are three questions we may ask about this temporal integration: Why does it occur? What are its limits? And when does the listener reach a decision about what he has heard? We will consider these questions in turn.

Why does temporal integration occur? We have seen that cues as diverse and as widely distributed as (1) the spectral and temporal properties of the vocalic portion preceding the silence, (2) the silence duration itself and (3) the spectral and temporal properties of the noise portion following the silence are all integrated into a unitary phonetic percept. Can we explain such integration on a purely auditory basis? Auditory integration does occur--for example, it is responsible for the perceptual coherence of homogeneous events such as the fricative noise--and surely we have much more to learn about such integration, especially in the case of complex acoustic signals. However, it seems to us quite implausible to suppose that purely auditory principles could ever account for perceptual integration of acoustic cues as heterogeneous and temporally spread as those we have dealt with here.

We encounter similar problems when we seek to explain our results in terms of feature detectors, as they have been postulated by several contemporary theorists (for example, Eimas and Corbit, 1973; Miller, 1977; Blumstein, Stevens, and Nigro, 1977). Consider again the case where the perception of a syllable-final stop consonant (GREAT vs. GRAY) depends on whether the fricative noise following the silence extends beyond a certain duration. If a single phonetic feature detector were responsible for the syllable-final stop, then its integrative power and complexity would have to be so great as to remove from the concept of feature detector the simplicity that is its chief attraction. Alternatively, there might be many simple auditory feature detectors, each responsive to elementary properties of the signal, whose outputs are integrated by a higher-level phonetic decision mechanism (see Massaro and Cohen, 1977). But that view fails to provide any principled reason why the outputs of certain feature detectors feed into a single phonetic decision in the way they do. Without reference to the articulatory system that produced the speech signal, the rules by which the detector outputs might be integrated would seem entirely arbitrary.

As we pointed out in the Introduction, we believe that the guiding principle of temporal integration in phonetic perception is to be found in the articulatory act that underlies the production of the relevant phonetic segment. By an "articulatory act" we mean, not a particular articulatory gesture, but all articulatory maneuvers that result from the speaker's "intention" to produce a given segment (such as a stop consonant). Thus, our definition of the articulatory act is intimately tied to the hypothesis that units of phone size are physiologically real at some early level in speech production. At the later articulatory level, we can distinguish individual gestures (such as closing and opening the jaw, raising the tongue tip, etc.) that form the components of the articulatory act. It is, of course, these several gestures that produce the several (and sometimes even more numerous) acoustic cues. The perceptual process by which the acoustic cues are integrated into a unitary phonetic percept somehow recaptures the gestures and also mirrors the processes by which they unfolded from a unitary phonetic intention (or motor program). We find it plausible to suppose that speech perception, as a unique biological capacity, has in fact evolved to reflect the equally species-specific capacity for speech production. The consequence is that, in a very real sense, the listener perceives directly the speaker's "intent"--the phonetically significant articulatory act. (For related views, see Fowler, 1977; Bailey and Summerfield, 1978; Summerfield, footnote 3.)

We turn now to our second question--about the limits of temporal integration. From the data of our experiments, we obtained an estimate according to the following considerations. The boundary between GRAY CHIP and GREAT CHIP indicates the maximal time over which the stop manner cues preceding the silence are still integrated with the cues following the silence into a single stoplike percept (affricate). Although the exact temporal interval varied with the strength of the stop manner cues preceding the silence (see Figure 6), a silence duration of 100 msec is a reasonably typical value. To this must be added the approximate temporal extent of the relevant cues preceding and following the silence--at least 100 msec for the duration of the vocalic portion and the fricative noise, respectively. We thus arrive at a temporal range of 300-350 msec for the integration of stop manner cues. This estimate is in good agreement with results on the single-geminate distinction for intervocalic stop consonants, since, as Pickett and Decker (1960) and Repp (1976) have shown, that boundary occurs around 200 msec of silence at normal rates of speech. Inasmuch as the manner cues following the closure interval (the formant transitions of the second vocalic portion) are shorter in this case (perhaps 50 msec), we arrive again at an integration period of about 350 msec. This coincidence is not surprising, since the articulatory gesture underlying an intervocalic stop consonant is similar to that for a stop consonant embedded between a vowel and a fricative. In our view, the range of temporal integration in perception reflects, not an auditory limitation--such as the duration of a preperceptual auditory store (Massaro, 1975)--but the maximal acceptable duration of the underlying articulatory act. Different articulatory acts may well be associated in perception with different ranges of temporal integration.

We thus arrive at our third question: When does the listener decide what he has heard? Before we can answer that question, we must point out that there are two logically distinct decisions the listener must make: (1) What phoneme has occurred? (2) Where does it belong? Thus, in the case of the GREAT SHIP--GRAY CHIP distinction, the listener must decide first that a stop consonant has occurred and, then, whether it belongs with the first or the second syllable. We see three possibilities for the temporal organization of the listener's decisions: (1) both the What and Where decisions occur after all relevant cues have been integrated; (2) the What decision occurs as soon as sufficient cues are available, but the Where decision is delayed until the end of the integration period; (3) both a What decision and a Where (default) decision are made as soon as sufficient cues are available, but the Where decision may be revised in the light of later information. We will discuss these hypotheses in turn.

The first hypothesis implies, in the case of GREAT SHIP, that the listener does not know whether he has heard a stop consonant until he has processed at least the first 120 msec of the fricative noise. This seems implausible on intuitive grounds. More likely, phonetic information accumulates continuously from the speech signal, and What decisions can be made, in principle at least, before all cues have been processed (also see Remington, 1977; Repp, 1976). If this were not so, we would have to assume that the relevant cues are integrated at a prephonetic level and thus are held in a temporary auditory memory--precisely the argument that we do not wish to make. On the other hand, if temporally separate cues (such as those preceding and following the silence in GRAY CHIP) are immediately translated into phonetic representations, temporal integration merely combines identical phonetic codes within a certain time span and thus is not dependent on auditory limitations. In terms of our experiment, this means that the listener already "knew" at the end of the vocalic portion of GRAY (which, as the reader may remember, contained sufficient stop maner cues to be perceived as GREAT in isolation) that a stop had occurred; the silence duration cue (if less than about 100 msec) and the noise duration cue (if less than about 120 msec) merely confirmed this perceptual knowledge.

The remaining two hypotheses differ in their assumptions about when the Where decision occurs. According to one hypothesis, the listener does not know whether he has heard GRAY or GREAT until he has processed the fricative noise; in other words, the Where decision is postponed until all relevant cues have been integrated. The alternative hypothesis assumes that the listener groups the stop consonant automatically with the preceding syllable, until later information leads him to revise that decision. This leads to the paradoxical prediction that, in an utterance heard as GRAY CHIP, the listener actually perceives GREAT for the brief moment that extends from the end of the vocalic portion to the end of the fricative noise, as he would have if CHIP had never occurred. This prediction can be tested empirically. Thus, Repp (1976) has shown in a reaction time experiment that practiced listeners can decide the identity of intervocalic stop consonants on the basis of the preclosure formant transitions alone. This supports the hypothesis that the What decision is made as soon as sufficient cues are available. The same procedure, if applied to GRAY CHIP and GREAT SHIP stimuli differing only in noise duration, would presumably reveal that the cues in GRAY (GREAT) are sufficient for deciding that a T has occurred. In other words, reaction times would be the same for GRAY CHIP and GREAT SHIP and (in practiced subjects) fast enough to indicate that the responses were indeed based on information preceding the silence. In another condition, the subjects might be asked to respond whenever they hear GREAT, rather than just T. If a similar result is obtained--namely, equally fast positive responses to GRAY CHIP and GREAT SHIP stimuli that differ in noise duration--we should find support for the hypothesis that a default Where decision accompanies every What decision. We hope to conduct such an experiment in the near future.

REFERENCES

Ainsworth, W. A. (1974) The influence of precursive sequences on the perception of synthesized vowels. *Language and Speech 17*, 103-109.

Bailey, P. J. and A. Q. Summerfield. (1978) Some observations on the perception of [s]+stop clusters. *Haskins Laboratories Status Report on Speech Research* SR-53, vol. 1,

Bastian, J. (1959) Silent intervals as closure cues in the perception of stop consonants. *Haskins Laboratories Quarterly Progress Report 33* (Appendix 1).

Bastian, J. (1960) Silent intervals and formant transitions as cues in the perception of stop phonemes. *Haskins Laboratories Quarterly Progress Report 35* (Appendix 2).

Bastian, J., P. D. Eimas and A. M. Liberman. (1961) Identification and discrimination of a phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America 33(A)*, 842.

Blumstein, S. E., K. M. Stevens and G. N. Nigro. (1977) Property detectors for bursts and transitions in speech perception. *Journal of the Acoustical Society of America 61*, 1301-1313.

Christie, W. M., Jr. (1974) Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America 55*, 819-821.

Cutting, J. E. and B. S. Rosner. (1974) Categories and boundaries in speech and music. *Perception and Psychophysics 16*, 564-570.

Darwin, C. J. and S. Brady. (1975) Voicing and juncture in stop-lateral clusters. *Journal of the Acoustical Society of America 57(A)* (Supp. No. 1), S24-S25.

Dorman, M. F., L. J. Raphael and A. M. Liberman. (1976) Further observations on the role of silence in the perception of stop consonants. *Haskins Laboratories Status Report on Speech Research SR-48*, 199-207.

Eimas, P. D. and J. D. Corbit. (1973) Selective adaptation of linguistic feature detectors. *Cognitive Psychology 4*, 99-109.

Finney, D. J. (1971) *Probit Analysis*. (Cambridge: Cambridge University Press).

Fowler, C. A. (1977) Timing control in speech production. Unpublished doctoral dissertation, University of Connecticut.

Fujisaki, H., K. Nakamura and T. Imoto. (1975) Auditory perception of duration of speech and non-speech stimuli. In *Auditory Analysis and Perception of Speech*, ed. by G. Fant and M. A. A. Tatham. (London: Academic Press), 197-220.

Gay, T. (1978) Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America 63*, 223-230.

Gerstman, L. (1957) Cues for distinguishing among fricatives, affricates, and stop consonants. Unpublished Ph.D. dissertation, New York University.

Kozhevnikov, V. A. and L. A. Chistovich. (1965) *Speech, Articulation, and Perception.* [U.S. Dept. of Commerce: National Technical Information Service (JPRS-305430)].

Lehiste, I. (1960) An acoustic-phonetic study of internal open juncture. *Phonetica Supplement 5*.

Massaro, D. W. (1975) Preperceptual images, processing time, and perceptual units in speech perception. In *Understanding Language. An Information-Processing Analysis of Speech Perception, Reading, and Psycholinguistics*, ed. by D. W. Massaro. (New York: Academic Press), 125-150.

Massaro, D. W. and M. M. Cohen. (1977) Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction. *Perception and Psychophysics 22*, 373-382.

Miller, J. L. (1977) Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America 62*, 641-648.

Nakatani, L. H. and K. D. Dukes. (1977) Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America 62*, 714-719.

Pickett, J. M. and L. R. Decker. (1960) Time factors in the perception of a double consonant. *Language and Speech 3*, 11-17.

Remington, R. (1977) Processing of phonemes in speech: A speed-accuracy study. *Journal of the Acoustical Society of America 62*, 1279-1290.

Repp, B. H. (1976) Perception of implosive transitions in VCV utterences. *Haskins Laboratories Status Report on Speech Research SR-48*, 209-233.

Diachronic Tone Splits and Voicing Shifts in Thai: Some Perceptual Data*

Arthur S. Abramson[†] and Donna M. Erickson

ABSTRACT

Proto-Tai is said to have had three phonemic tones and four consonantal voicing categories, which would have been inherited by Old Thai (Siamese). Correlations between tones and initial consonants across the Tai languages have led to the positing of tonal splits conditioned by the voicing states of initial consonants with a subsequent shifting of voicing features in certain lexical classes. Thus for each tone of Old Thai, words with initial voiced consonants developed a lower tone and words with initial voiceless consonants, a higher tone. Two types of experiment were designed to test the phonetic plausibility of the argument: (1) CV syllables were synthesized with three values of voice onset time (VOT) acceptable as Thai /b p ph/. Each of these was combined with a continuum of F_0 contours that had previously been divided perceptually into the high, mid and low tones. These syllables were played to native speakers of Thai for tonal identification. (2) Labial stops with nine values of VOT separable into /b p ph/ categories were coupled on synthetic mid-tone and low-tone CV syllables with upward and downward F_0 onsets varying in extent and duration. The resulting syllables were played to native speakers for identification of the initial consonants. The historical argument receives some support from the experimental data.

The term tonogenesis, apparently first used by James Matisoff (1970), can mean the emergence of phonologically distinctive tones in a previously toneless language under the influence of certain contextual features. Although in a given case the best historical reconstruction may lead to just that conclusion, we see no reason to believe that tonal distinctions should not be just as primitive as vocalic or consonantal distinctions in a protolanguage. A further use of the term tonogenesis has been as a label for the splitting of old tonal categories into a larger number of tones. A consensus of historical linguists is that such has been the development in the Tai family of languages. J. Marvin Brown (1975) refers to the "great tone split ... that swept through China and northern Southeast Asia nearly a thousand years ago."

---

*This paper was presented before the Linguistic Society of America, Honolulu, August 1977.

[†] Also the University of Connecticut.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 2 (1978)]

During the period of the emergence of its daughter languages, Proto-Tai is generally said to have had three phonemic tones on "smooth" syllables (those ending in vowels, glides or nasals) and four voicing categories for initial consonants, which would all have been inherited by Thai (Siamese). With some help from the ancient writing systems, examination of correlations between tones and initial consonants has led to the positing of tonal splits conditioned by the voicing states of initial consonants with a subsequent phonological shifting of voicing features in certain lexical classes [Maspero, 1911; Li, 1947, 1977; Haudricourt, 1956; (Gedney, see footnote 1]. This development purportedly underlies the system of five tones and three consonantal voicing categories of modern Thai. Thus, ignoring the special problems of one of the four classes of consonants, the so-called glottalized consonants, we find that for each tonal category of Old Thai, words with initial voiced consonants developed a lower tone, and words with initial voiceless consonants, a higher one. Thus the three original tones would have split into six. In fact, given the vicissitudes of language change spread over related languages, we find that Central Thai, which is the dialect of the Bangkok region and the official language of Thailand, has only five tones, while other dialects and other Tai languages have six or more, with differences among them in tonal shapes as well.

Independently of these historical hypotheses, it has been known for some time that the fundamental frequency of a syllable beginning with a voiced consonant is likely to be lower than that of a syllable beginning with a voiceless consonant (House and Fairbanks, 1953; Lehiste and Peterson, 1961). We know that for Thai (Gandour, 1974; Erickson, 1975) and other languages (Hombert, 1975), voiced initials are in fact accompanied by an upward movement of fundamental frequency, and voiceless consonants, by a downward movement; both of these perturbations then tend to move back toward the prosodic norm of the syllable as a whole. While the physiological mechanisms underlying these perturbations are still rather controversial, fundamental-frequency movements of comparable magnitude have been shown by Hombert to be quite perceptible. It has also been found that either in exaggerated form (Haggard, Ambler and Callow, 1970) or within more or less normal ranges (Fujimura, 1971; Abramson, 1974), such perturbations can influence phonemic judgments of voicing. If we assume these findings in production and perception to be universal and thus to apply to Old Thai, we might suppose that speakers of the language, already accustomed to a three-way tonal contrast, were psychologically receptive to the pitch fluctuations normally occurring with voicing distinctions. However, attention was gradually shifted to syllable nuclei as pitch perturbations were more and more enhanced in perception and production, and away from syllable initials, to the detriment of the latter. That is, the phonemicization of the pitch fluctuations, yielding an increase in tonal categories, helped to keep the old lexical classes apart even while the consonantal voicing categories, to some extent under the influence of pitch, decayed, shifted and even coalesced.

---

[1] Gedney, W. J. Future directions in comparative Tai linguistics. (unpublished manuscript).

Our aim was to examine the perceptual plausibility of the foregoing historical arguments. We approached the problem in two ways. First, we tested for the effects of systematic pitch perturbations on the identification of initial stop consonants. Then we experimented with the effects of the voicing states of initial stop consonants on tone identification. In order to have incremental control over the phonetic dimensions of interest to us, we followed the common practice in experimental phonetics of using synthetic speech.

Techniques of acoustic analysis and synthesis have shown that the voiced, voiceless unaspirated, and voiceless aspirated stops of Modern Thai lie along a dimension of voice onset time (VOT), namely, the temporal relation between the closing of the glottis for audible pulsing and the release of the occlusion of the initial stop (Lisker and Abramson, 1964; Abramson and Lisker, 1965). VOT itself is simply an instance in utterance-initial position of a more general phenomenon of laryngeal timing in consonant distinctions (Abramson, 1977). In Figure 1 are plotted the responses of 48 native speakers of Thai to a synthetic continuum of variants of VOT in labial stop consonants with the vowel /aa/ on the mid tone. The abscissa shows values of VOT, with voicing lead in negative numbers and voicing lag in positive numbers, while zero means voice onset at the moment of release. For voicing lead, low-frequency harmonics are present before the release during the simulated occlusion; for voicing lag, until the moment of pulsing onset after the release, the upper formants are filled with noise to simulate aspiration, and the first formant is omitted. The ordinate shows percent identification of each of the stops. As can be seen, the three expected voicing categories emerge.

Having demonstrated the sufficiency of VOT as a cue to the three-way voicing distinction in Thai, we turned to the question of the effect of pitch perturbations at syllable beginnings on consonant identification. The stimuli were made for this experiment by varying VOT and extent and duration of initial fundamental-frequency shifts. The basic syllable pattern for all stimuli was a set of formant transitions appropriate to the labial place of articulation and steady-state formants for a vowel of the type [aː]. With the data from Figure 1 as a baseline, we chose nine VOT values ranging from -100 to +80 msec to span the three voicing categories. An acceptable mid tone was produced by using a level fundamental frequency at 120 Hz with a small drop at the end. As shown in Figure 2, four shifts of fundamental frequency were applied to the beginning of the contour, with falls from 10 Hz and 20 Hz above and rises from 10 Hz and 20 Hz below. These values were derived from production data published by Erickson (1974). To these four was added a fifth variant with a level onset, that is, no shift. Finally, because of some uncertainty in the literature as to the appropriate duration of such shifts, we synthesized them with three time spans: 50, 100, and 150 msec. The resulting 117 stimuli were presented in a number of randomizations for identification as to voicing category. The results for 46 speakers of Thai for the 100-msec condition are presented in Figure 3. From top to bottom the three graphs show identifications of the stimuli as /b/, /p/, and /ph/, respectively. As shown by the coded lines, it is the boundary between the voiced and the voiceless unaspirated stops along the VOT dimension that is much affected by fundamental-frequency perturbations, while the boundary between the voiceless unaspirated stop and the voiceless aspirated stop hardly

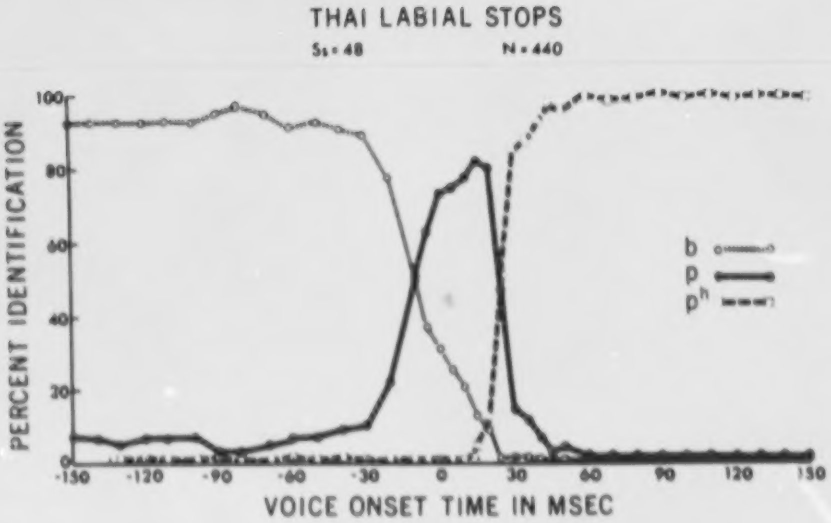


Figure 1: Thai identifications of synthetic labial stops varying in voice onset time (VOT).

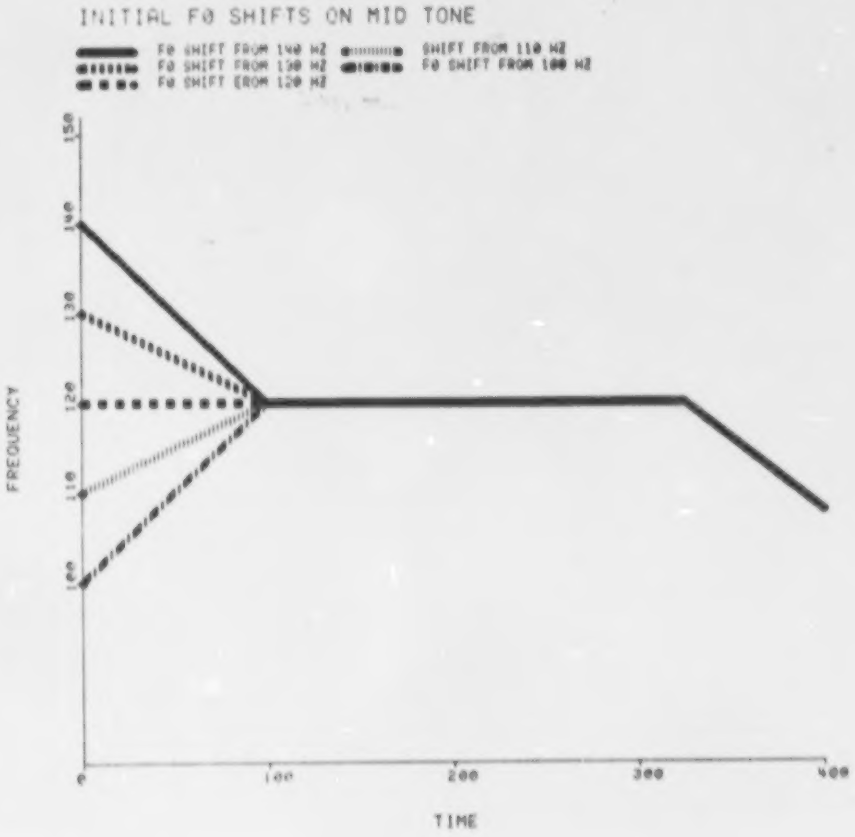


Figure 2: Initial F_0 shifts for stop-voicing identification on the mid tone.

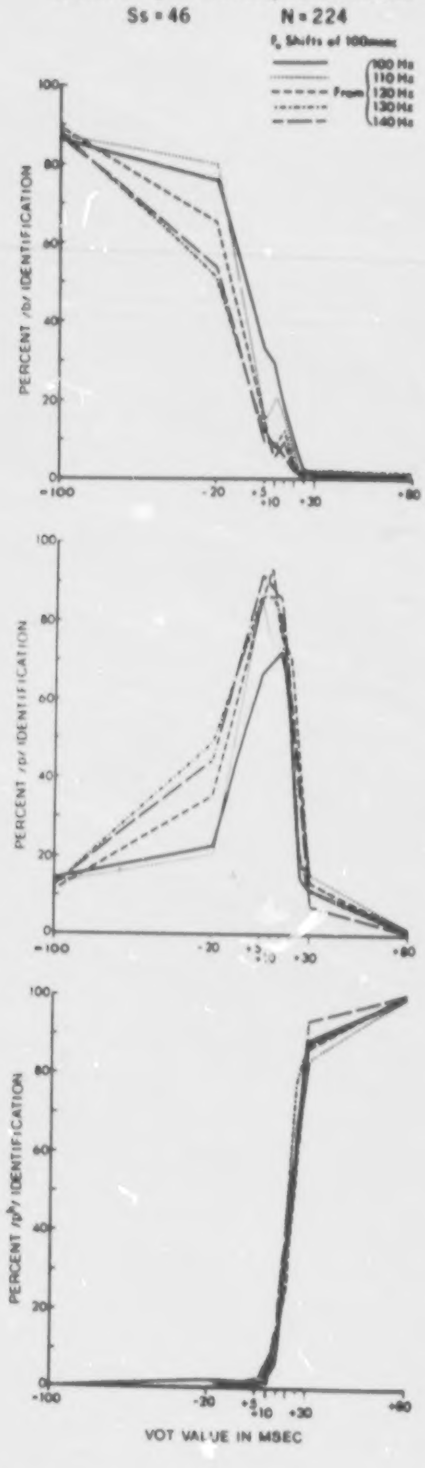


Figure 3: Effects of initial F_0 shifts on the identification of voicing in Thai initial stops.

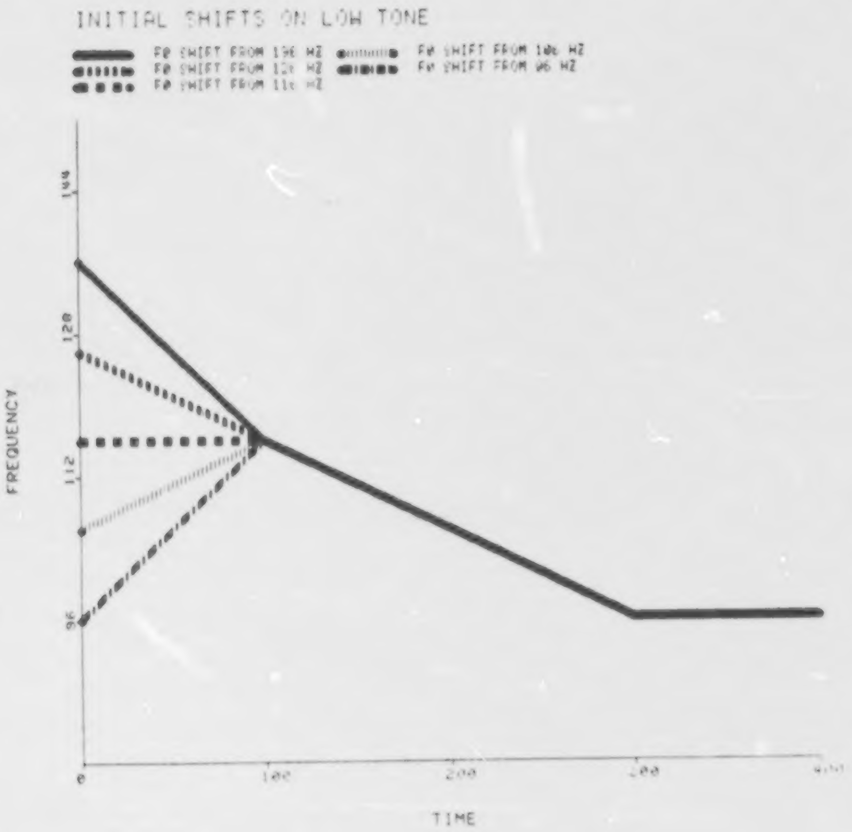


Figure 4: Initial F_0 shifts for stop-voicing identifications on the low tone.

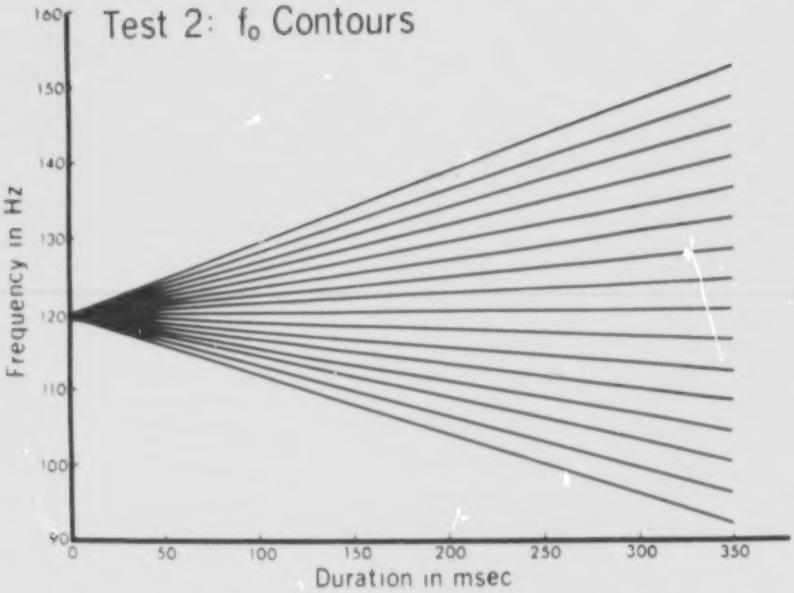


Figure 5: A set of F_0 contours yielding identifications as high, mid and low tones.

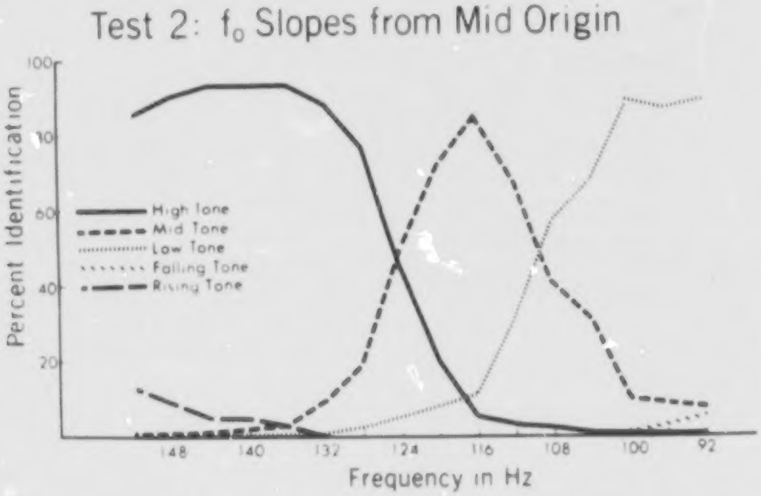
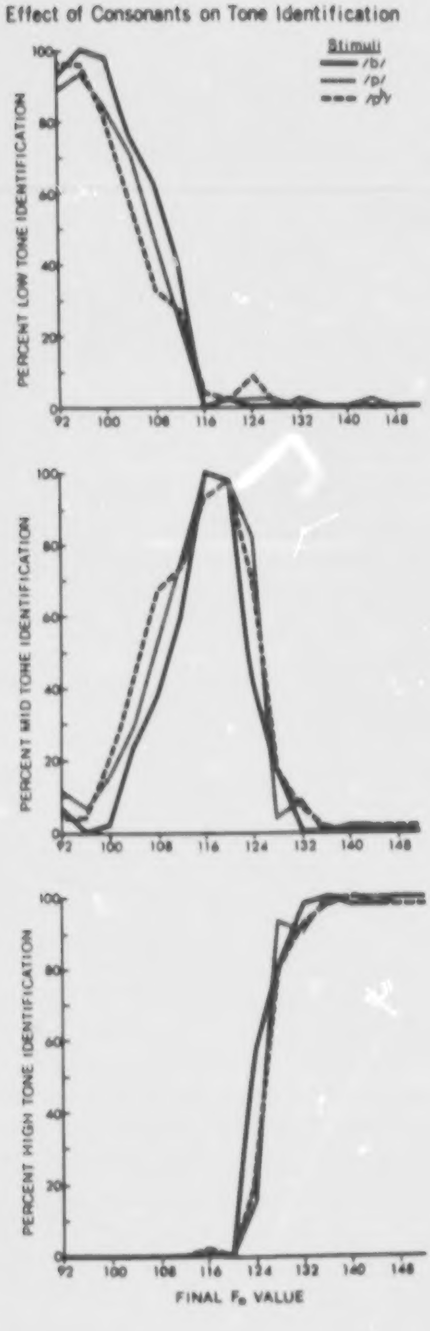


Figure 6: Identification functions for the contours of Figure 5.



Effect of Consonants on Tone Identification

Figure 7: Effects of the voicing states of initial consonants on the tonal identification of the contours in Figure 5.

varies. An analysis of variance shows these effects to be highly significant. The overall differences between the three time spans were not significant. It is for this reason that only one duration is given in Figure 3.

Thus it is clear that at least one voicing boundary can be pushed about by perturbations of fundamental frequency, but it is important to know that this phenomenon is not restricted to the mid tone used up to this point. Reasoning that three reconstructed tones of Proto-Tai might well have been phonetic approximations to the relatively static high, mid, and low tones of Central Thai, we chose the low tone for the next experiment. For this experiment a fundamental-frequency contour that had been shown to be acceptable for the low tone (Abramson, 1962, 1975) was imposed on the same syllable type; it began at 116 Hz and dropped to 96 Hz where it leveled off until the end of the syllable. Except for a limit on the duration of each perturbation to 100 msec, the stimuli for this experiment were analogous to those made on the mid tone. That is, as shown in Figure 4, there were downward shifts of fundamental frequency from 10 and 20 Hz above the starting point of the basic tonal contour and upward shifts from 10 and 20 Hz below, in addition to one unperturbed contour. The responses of eight subjects to randomizations of these stimuli are essentially the same as those in the experiment on the mid tone. The great discrepancy between numbers of subjects available for the two experiments makes it very difficult to do a detailed statistical comparison across the two tonal conditions; nevertheless, the main effects are repeated; it is only the boundary between the voiced stop and the voiceless inaspirate that is affected. An extension of this research will be to try the same experiment with the high tone. Will it show the same pattern of responses? Since, in other studies the mid and low tones have been shown to be confusable under certain conditions (Abramson, 1976a), but never the high tone with either of them, it is conceivable that the high tone provides a suitable context for accompanying downward perturbations of its onset to have more of an effect on stop identification.

The foregoing data raise a question: Why is the boundary between the two voiceless categories not affected? A reasonable explanation may be that once one reached far enough into the voicing-lag part of the VOT dimension, the resulting stimuli are psychoacoustically very different from variants with lower VOT values--as very audible noise-excitation of the formants is present in the signal. We think that this aspiration noise is so powerful a cue that any accompanying pitch shifts, even if audible, cannot affect labeling responses. The lack of effect of pitch perturbation on the voiceless aspirate in our data accords well with the historical observation that the voiceless aspirate of the protolanguage has persisted into Modern Thai. The shift of the voiced stop of the protolanguage to modern voiceless aspirate, however, seems to require a more indirect explanation. In this connection, it is important to note that in most other Tai languages this phoneme became a voiceless inaspirate.

Thus we see from the preceding two experiments that the boundary between voiced stops and voiceless inaspirates is affected by initial fundamental-frequency perturbations. To continue our research into the question of interactions between voicing states and pitch in the history of Thai, we turned to an experiment on the effects of initial stop consonants on the identification of tones. As shown in Figure 5, we used a fan-shaped continuum of fundamental-frequency variants with a common origin that had been shown previously (Abramson, 1976b) to be perceptually divisible into the three static tones, high, mid, and low. The tonal variants all started at 120 Hz and moved to the end points ranging from 152 to 92 Hz in four-Hz steps. The labeling responses of 31 Thai subjects to this continuum in that study are shown in Figure 6. Here we can see that they essentially divided the continuum into three tones. We synthesized syllables with VOT values appropriate to /baa paa phaa/ with all 16 tonal variants. Randomized lists of the stimuli were presented to Thai subjects for identification of both the stops and tones.

Figure 7 shows the effects of the initial stop consonants on tone identification. The graphs from top to bottom give the results for the identification of the low, mid, and high tones, respectively, as they are affected by the three voicing categories of stops. The three stops are indicated by the coded lines. Essentially, the data show that the tone identification is affected by the consonant categories, but in a somewhat paradoxical way. For the low-tone identification function, the voiced stop entails a significantly higher number of low-tone judgments than does the voiceless aspirated stop. For the high-tone identification function, the voiced stop entails a significantly higher number of high-tone judgments than do both the voiceless stops.

The paradox is the direction of the boundary shift for the voiced stop. It is at a higher fundamental frequency for the boundary between the mid and low tones, and at a lower fundamental frequency for the boundary between the mid and high tones. The best interpretation we have to offer at this time is the following. In the case of the low tone, we would expect that since the Thai speaker associates lower pitch with voiced initial consonants, he does not need as low a value of fundamental frequency to hear a low tone on the syllable beginning with /b/. In the case of the high tone, our reasoning is somewhat the opposite. That is, with the high tone, again the listener associates an inherent lowness with the /b/ consonant, but in this case, he compensates for the expected lowness by allowing syllables beginning with /b/ to be heard as high tones at a lower frequency range. This kind of interpretation assumes a difference between low and high tones in perceptual processing yet to be understood.

To summarize this experiment, we can say that the tonal boundaries are affected by the stop categories, although we do not yet understand the exact reason for the nature of this interaction. A more complicated experiment planned for the future is to combine the two approaches used here, namely to perturb the onsets of the 16 fundamental frequency variants in association with the voicing characteristics of the initial stops.

In conclusion, then, our experiments show that fundamental-frequency perturbations affect consonant categories and that consonant categories affect tone labeling of fundamental-frequency continua. Thus, our data lead us to the conclusion that, by and large, the historical arguments concerning interactions between tone splits and voicing shifts are perceptually plausible. As pitch perturbations loomed larger in the consciousness of the speakers and gradually took on phonemic status, one might expect that the voicing states of initial consonants would have been reassessed perceptually and rearticulated to furnish new production norms, thus helping to bring about shifts in tone and consonant categories.

REFERENCES

Abramson, A. S. (1962) The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments. (Bloomington: Indiana University Research Center in Anthropology, Folklore and Linguistics, Pub. 20) [International Journal of American Linguistics 28 (2, pt. 3)].

Abramson, A. S. (1974) Pitch in the perception of voicing states in Thai: Diachronic implications. Linguistics Society of Meeting Handbook, Forty-Ninth Annual Meeting. (A), 1. [Also in Haskins Laboratories Status Report on Speech Research SR-41, 165-174].

Abramson, A. S. (1975) The tones of Central Thai: Some perceptual experiments. In Studies in Tai Linguistics in Honor of William J. Gedney, ed. by J. G. Harris and J. R. Chamberlain. (Bangkok: Central Institute of English Language), 1-16.

Abramson, A. S. (1976a) Thai tones as a reference system. In Tai Linguistics in Honor of Fang-Kuei Li, ed. by T. W. Gething, J. G. Harris and P. Kullavanijaya. (Bangkok: Chulalongkorn University Press), 1-12.

Abramson, A. S. (1976b) Static and dynamic acoustic cues in distinctive tones. Journal of the Acoustical Society of America 59, S42(A). [Also in Haskins Laboratories Status Report on Speech Research SR-47, 121-127.

Abramson, A. S. (1977) Laryngeal timing in consonant distinctions. Phonetica 34, 295-303.

Abramson, A. S. and L. Lisker. (1965) Voice onset time in stop consonants: Acoustic analysis and synthesis. In Proceedings of the Fifth International Congress of Acoustics, ed. by D. E. Commins, A51 (Liege: G. Thone).

Brown, J. M. (1975) The great tone split: Did it work in two opposite ways? In Studies in Tai Linguistics in Honor of William J. Gedney, ed. by J. G. Harris and J. R. Chamberlain. (Bangkok: Central Institute of English Language), 33-48.

Erickson, Donna. (1974) Fundamental frequency contours of the tones of Standard Thai. Pasaa 4, 1-25.

Erickson, Donna. (1975) Phonetic implications for an historical account of tonogenesis in Thai. In Studies in Tai Linguistics in Honor of William J. Gedney, ed. by J. G. Harris and J. R. Chamberlain. (Bangkok: Central Institute of English Language), 100-111.

Fujimura, O. (1971) Remarks on stop consonants: Synthesis experiments and acoustic cues. In Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen, ed. by L. L. Hammerich, R. Jakobson and E. Zwirner. (Copenhagen: Akademisk Forlag), 221-232.

Gandour, J. (1974) Consonant types and tones in Siamese. Journal of Phonetics 2, 337-350.

Haggard, M., S. Ambler, and M. Callow. (1970) Pitch as a voicing cue. Journal of the Acoustical Society of America 47, 613-617.

Haudricourt, A. G. (1956) De la restitution des initiales dans les langues monosyllabiques: le probleme du Thai commun. Bulletin de la Societe de Linguistique de Paris 52, 307-322.

Hombert, J. M. (1975) Towards a theory of tonogenesis: An empirical, physiologically and perceptually-based account of the development of tonal contrasts in languages. Doctoral dissertation. (Berkeley: University of California).

House, A. S. and C. Fairbanks. (1953) The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America 25*, 105-113.

Lehiste, I. and G. E. Peterson. (1961) Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America 33*, 419-425.

Li, Fang-Kuei. (1947) The hypothesis of a pre-glottalized series of consonants in primitive Tai. *Academia Sinica 11*, 177-188.

Li, Fang-Kuei. (1977) *A Handbook of Comparative Tai*. (Honolulu: University Press of Hawaii).

Lisker, L. and A. S. Abramson. (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word 20*, 384-422.

Maspero, H. (1911) Contribution à l'étude du système phonétique des langues Thai. *Bulletin de l'École Française d'Extrême-Orient 19*, 152-169.

Matisoff, J. (1970) Glottal dissimilation and the Lahu high-rising tone: a tonogenetic case-study. *Journal of the American Oriental Society 90*, 13-44.

A Range Effect in the Perception of Voicing*

Susan A. Brady† and Christopher J. Darwin††

ABSTRACT

The location of the voicing boundary in the perception of initial stop consonants is shown to vary according to the range of voice onset times used in a block of trials and according to the order in which blocks covering different ranges are presented. Although these range effects introduce methodological complications into the interpretation of adaptation experiments, they appear to be qualitatively different from adaptation effects and, it is suggested, may provide a metric for assessing the auditory tolerance of phonological categories.

INTRODUCTION

The numerical categories that subjects assign to particular stimuli along arbitrary dimensions, such as circle size or line length, are influenced by both the range and relative frequency of occurrence of stimulus values in the experiment (Helson, 1964). A particular stimulus will, for example, be assigned to a lower category when the range of stimuli used extends further to the high category end. The effects of frequency are more complicated but suggest a tendency for subjects to place equal numbers of stimuli in each category. Frequently occurring adjoining stimuli thus are allocated an unduly wide range of categories (Parducci, 1974).

These effects of range and frequency are commonly found for dimensions that do not fall into "natural categories" (Rosch, 1973) and the question has been raised (Sawusch and Pisoni, 1974; Studdert-Kennedy, 1976) whether the perception of category boundaries in speech is immune from such effects. Changes in the relative frequency with which different stops, taken from either a voicing (Sawusch and Pisoni, 1974) or a place of articulation (Sawusch, Pisoni and Cutting, 1974) continuum, are played to subjects do not influence the position of their phoneme boundaries. On the other hand,

---

*In press in the Journal of the Acoustical Society of America.

†University of St. Andrews, Fife, Scotland.

††University of Sussex, Brighton, U.K.

Acknowledgment: Experiment I was supported by a grant from NICHD to Haskins Laboratories and Experiment II by a grant from the British SRC to Christopher J. Darwin.

similar frequency manipulations for fundamental frequency do move the point at which subjects change from using the label "high" to using "low" (Sawusch and Pisoni, 1974; Sawusch et al., 1974).

This suggested immunity is of considerable theoretical interest and potentially removes a number of methodological problems (Poulton, 1973) from the interpretation of, for example, adaptation experiments, whose results have generally been interpreted differently from range and frequency effects. However, it has been pointed out that range and frequency effects operate independently (Parducci, 1974), so it is possible that speech categories might be prone to range effects while being immune from frequency effects.

In the two experiments presented here, we first show that small but significant changes in the voicing boundary for stops can be obtained when the range of sounds presented within a block of trials is varied. We then go on to discuss the significance of this range effect in adaptation experiments and also to raise the possibility that range effects may be used to map the auditory tolerance of our internal phonological categories.

EXPERIMENTAL DESIGN AND PROCEDURE

In both experiments a synthetic voice onset time (VOT) continuum was used. An alveolar stop was synthesized on the Haskins Laboratories Parallel Formant Synthesizer before the diphthong /aI/ with 11 different VOT values from 5 to 55 msec in 5-msec steps, in order to give a continuum of sounds perceived as /daI/ with short VOT and /taI/ with long. The acoustic correlates of the change in VOT were a cut-back in first formant amplitude and a substitution of hiss for buzz excitation. In the first experiment each syllable was preceded by a carrier phrase "I may," but in the second experiment this was omitted. The two experiments were otherwise of identical design. Subjects listened to blocks of 40 trials in which the stimuli were drawn from five contiguous VOT steps: A(5-25 msec), B(15-35 msec), C(25-45 msec), D(35-55 msec). They also listened to an additional block covering the entire range of 11 steps: E(5-55 msec). Each block contained a random ordering of eight examples of each stimulus from the appropriate range. In each experiment, eight groups of four subjects listened to different orderings of these blocks; half the groups started with E and half finished with it, and within these halves each of the four groups took a different predecessor-balanced Latin-square order. The subjects, who for the first experiment were undergraduates at the University of Connecticut, and for the second were at the University of Sussex, were instructed to label each sound as being either more like a "d" or a "t." It was made clear to subjects that within a block of trials the proportion of either category could vary between zero and 100 percent. Subjects were encouraged to remember this and to judge each trial independently.

Results

Percent "d" responses for each VOT are shown in Figure 1 according to the range used in each block of trials, averaged across the four groups in each experiment who listened to the entire range block (E) prior to listening to the subset ranges (A, B, C, D) and also across those who listened to the E block last. There are significant differences between the responses given to

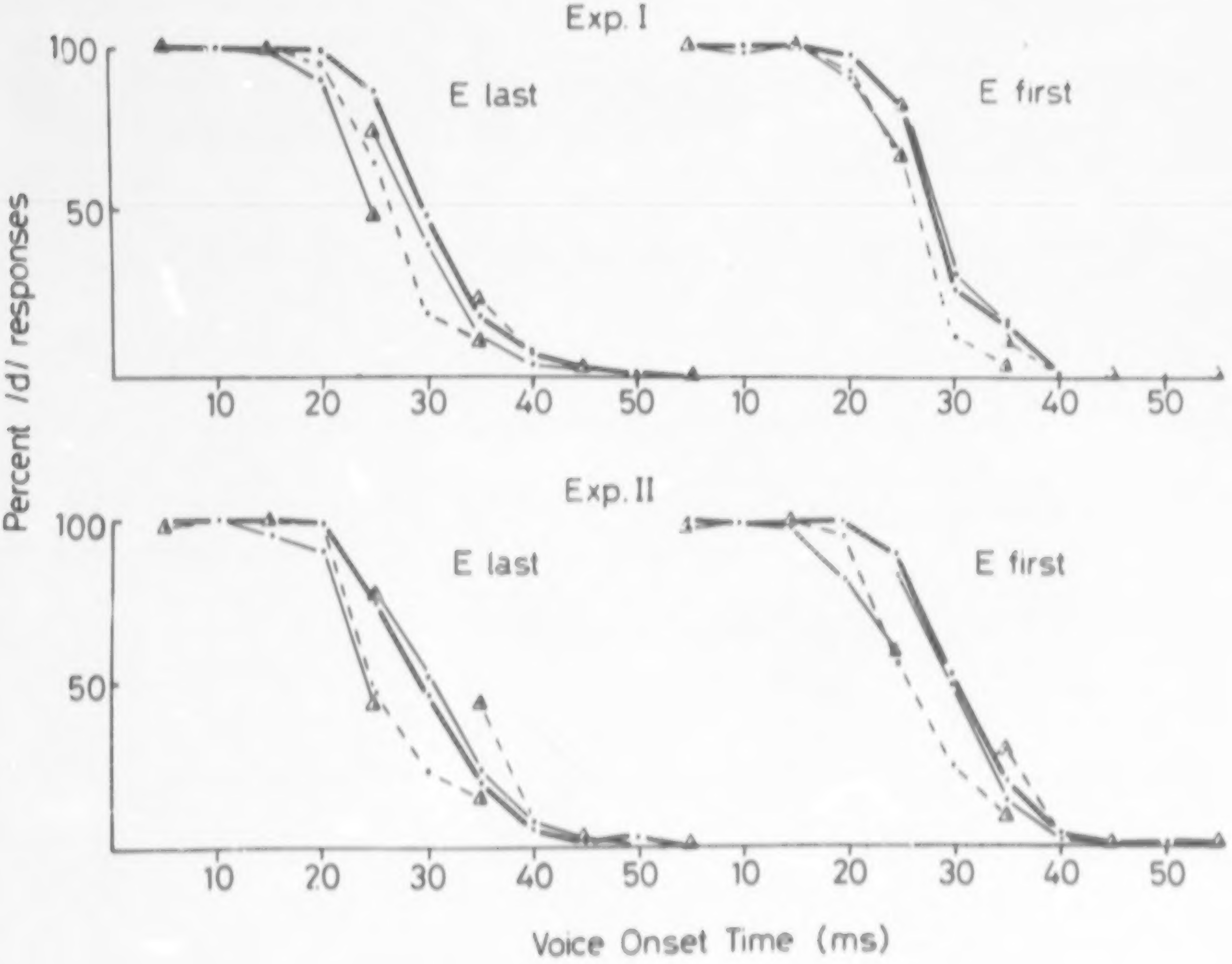


Figure 1: Percent "d" responses to stimuli differing in voice onset time between [daI] and [taI]. Each quadrant presents data from a different group of 16 subjects. Those in Experiment I heard the target syllable preceded by "I may ..."; those in Experiment II heard it in isolation. The two left-hand quadrants are for subjects who heard a block of trials covering the entire VOT range last, and the two right-hand ones for those woho heard it first. The data points for the entire range are connected by a wide line. Subjects also heard blocks of trials in which the stimuli were from subranges. Their responses to these are shown by narrow solid (ranges A and C) or narrow broken (B and D) lines. Triangles mark data points at the ends of the subranges.

a particular VOT, depending on the range in which it occurred. These differences are, of course, larger at VOT values near the boundary. Statistical analysis is problematic because of necessarily missing data, but the following tests have been performed:

(1) Since ranges B, C and E all cover the three crucial middle VOT values of 25, 30, 35 msec, an analysis of variance is possible on this restricted set of data. This showed a significant effect of range on the total number of voiced responses for each of the four sets of data illustrated in Figure 1, each with $p < 0.005$. Figure 1 shows that this effect is due to a particular VOT stimulus receiving fewer voiced responses when it occurs as part of a block of trials in which the range covers shorter VOT values, than when the range covers longer VOT values. There is also an interaction of range effect with the order in which the blocks were taken ($p < 0.005$ for all except Experiment II with E last, which gave $p < 0.05$).

(2) Looking at the three middle VOT values separately, Friedman two-way analyses of variance showed significant effects of range on total voiced responses at better than the 1 percent level for the 25 and 30 msec VOTs in each of the two experiments when the entire range (E) was last, and also at better than the 2 percent level in Experiment II for each of the three middle VOTs when range E was last. No significant variation was found for these three individual VOTs in Experiment I when range E came first.

(3) Any change in the pattern of results between the two experiments, which differed in whether the precursor "I may ..." was present, was assessed by two analyses of variance similar to those in (1), but with Experiment I versus II as an additional dimension. For subjects receiving the entire range (E) last, there was only a weak interaction of experiment number with range on the number of voiced responses ($p < 0.05$), but for subjects who took the entire range first, experiment number gave a three-way interaction with range and the order in which blocks were taken ($p < 0.005$).

Averaging over block orders, the range effects thus appear to be rather larger and more reliable when the block with the entire range is presented last than when it comes first, and there are also significant influences on the range effect of the order in which A, B, C and D were presented. These interactions indicate that the extent of the range effect is not restricted to a particular block, but rather that the position of the category boundary is influenced by the range of preceding blocks as well as that of the present block. Since the present design confounds a block's predecessor with its serial position, these effects cannot be analyzed systematically here, but a working hypothesis suggested by the data is that the effective range is determined by all the preceding blocks, perhaps weighted in favor of the current block.

Interpretation of the three-way interaction of experiment number (or presence of precursor) with range and block order, like the interpretation of the two-way interaction between range and block order, is complicated by the confounding of a block's serial position with its predecessor; however, inspection of the data suggests that it is mainly due to the sounds in range B. When there is a precursor, this range is heard as progressively more voiced the later in the experiment it is presented. When there is no precursor the same pattern occurs, except that when B is heard as the first block its sounds are consistently heard as more voiced than for later presentations.

There is thus some evidence that both a precursor and block order can influence the effect of stimulus range; but they are not sufficiently powerful to remove the effect, since we still find a significant (though reduced) range effect in the least favorable condition, when a precursor is present and the entire range is presented first.

Discussion

These two experiments give clear evidence that the perceived voicing of a sound depends quite markedly on the range of other sounds presented before it in an identification experiment. The more voiced the previous sounds are, the more voiceless they will appear. Similar effects of range are also present in the results of Lisker, Liberman, Erickson and Dechovitz [1975; Lisker, 1975, Figure 1]. There the VOT boundary in a /da/-/ta/ distinction was subject to the range of first formant transition durations used in the experiment. Again, their data indicated that the more voiced the companion stimuli, the more voiceless would a particular stimulus sound.

It might be argued that the results of our Experiment II could be interpreted as reflecting a general linguistic mechanism that compensates for the apparent rate of articulation of the speaker; this is unlikely because a constant rate precursor does not eliminate the effect (Experiment I) and because this hypothesis would not predict any range effect in Lisker's data. Linguistic categorizations, like other perceptual distinctions, rely on perceptual contrast and, it would seem, the decision mechanisms that register this contrast can be influenced by factors that are apparently linguistically irrelevant. Perhaps such flexibility should be welcomed in a communication system that must cope with the idiosyncrasies of individual speakers and with the varied distortions to which the speech signal is subjected.

The mean results shown in Figure 1 suggest that the range effects found here may be asymmetrical, with subjects being more willing to perceive as unvoiced a sound to the long-VOT end of a short-VOT range than to perceive as voiced a sound to the short-VOT end of the corresponding long-VOT range. An adequate statistical assessment of this is complicated by the interaction with block order and by the phoneme boundary not being exactly in the middle of the entire range, but if borne out in a more suitable experimental design, it would suggest that the extent to which range effects can influence phoneme boundaries is limited by the phonetic plausibility of the result. There is considerable variation of VOT in natural prestressed aspirated initial stops within the range of values that we have used for this experiment, and so range effects may be free to operate over these phonetically plausible values; but it is rare to find a natural apical unaspirated stop with a VOT of greater than about 20 msec before a stressed vowel, and perhaps this constrains the extent of range effects. This hypothesis of phonetic plausibility could perhaps also account for the much larger context effects noted by Eimas

(1962)[1] in labeling data for triads of vowels than for triads of stops differing in place of articulation. The extent of range effects may thus provide a metric for assessing the auditory limits of our internal phonological categories.

The range effect found here and evident in other experiments might also contribute to the shift in stop voicing boundary found in adaptation experiments following repeated presentation of either a voiced or a voiceless stop, or some of the acoustic cues serving the voicing distinction [Eimas and Corbit (1973); Eimas, Cooper and Corbit (1973); Miller and Eimas (1975); Ades (1976)]. In particular, the early trials in the test phase of an adaptation experiment might form, with the adapting sound, a range that would differ depending on the adaptor. Although we cannot rule out the possibility of some of the adaptation effect being due to range, it is unlikely that all of the adaptation effect can be thus explained. The reason for this is that while adaptation effects are larger following adaptation to a voiceless stop [Eimas and Corbit (1973); Eimas et al. (1973); Miller and Eimas (1975)], our range effect seems to be greater for ranges extending into the voiced end of the voicing continuum. This difference in asymmetry may reflect a difference in underlying mechanisms, with adaptation being attributable to a change in the auditory representation of acoustic events, through adaptation of complex auditory feature detectors and range effects being attributable to a changed phonemic interpretation of a constant auditory representation. The greater phoneme boundary shift following voiceless adaptation could then be due to the voiceless adaptor reducing the sensitivity of a detector for long VOTs, which is perceptually more salient than the detector for low frequency first formant onset that might be reduced in sensitivity following adaptation to a voiced stop (Ades, 1976). The greater shift in phoneme boundary for a predominantly voiced than for a predominantly voiceless range, on the other hand, could be explained by the hypothesis of phonetic plausibility described earlier.

Thus, although we have shown that whatever natural categories we might possess for the voicing distinction are not immune from the biasing effects of range, nevertheless the size of the changes we find is quite small and they perhaps reflect in their magnitude the bounds of these internal categories.

REFERENCES

Ades, A. E. (1976) Adapting the property detectors for speech perception. In *New Approaches to Language Mechanisms*, ed. by R. J. Wales and E. Walker. (Amsterdam: North-Holland).

Eimas, P. D., W. E. Cooper, and J. D. Corbit. (1973) Some properties of linguistic feature detectors. *Perception & Psychophysics* *13*, 247-252.

Eimas, P. D. and J. D. Corbit. (1973) Selective adaptation of linguistic feature detectors. *Cognitive Psychology* *4*, 99-109.

Helson, H. (1964) *Adaptation Level Theory*. (New York: Harper and Row).

Lisker, L. (1975) In (qualified) defense of VOT. *Haskins Laboratories Status Report on Speech Research SR-44*, 109-117.

---

[1] Eimas, P. D. (1962) A study of the relation between absolute identification and discrimination along selected sensory continua. Unpublished doctoral dissertation, University of Connecticut.

Report on Speech Research SR-44, 109-117.

Lisker, L., A. M. Liberman, D. M. Erikson, and D. Dechnovitz. (1975) On pushing the voice-onset time (VOT) boundary about. *Journal of the Acoustical Society of America 57*, S50(A).

Miller, J. A. and P. D. Eimas. (1975) Studies on the selective tuning of feature detectors for speech. *Developmental Studies of Speech Perception 3*, (Progress Report of the Walter J. Hunter Laboratory of Psychology, Brown University), 103-116.

Parducci, A. (1974) Contextual effects: A range-frequency analysis. In *Handbook of Perception*, vol. 2, ed. by E. C. Carterette and M. P. Friedman. (New York: Academic).

Poulton, E. C. (1973) Unwanted range effects from using within-subject experimental designs. *Psychological Bulletin 80*, 113-121.

Rosch, E. H. (1973) Natural categories. *Cognitive Psychology 4*, 328-350.

Sawusch, J. R. and D. B. Pisoni. (1974) Category boundaries for speech and non-speech sounds. *Journal of the Acoustical Society of America 55*, 436(A).

Sawusch, J. R., D. B. Pisoni, and J. E. Cutting. (1974) Category boundaries for linguistic and non-linguistic dimensions of the same stimuli. *Journal of the Acoustical Society of America 55*, S55(A).

Studdert-Kennedy, M. (1976) Speech Perception. In *Contemporary Issues in Experimental Phonetics*, ed. by N. J. Lass. (New York: Academic), 243-293.

A Note on Perceptuo-Motor Adaptation of Speech

Quentin Summerfield[†], Peter J. Bailey[††] and Donna Erickson

ABSTRACT

Existing data on perceptuo-motor adaptation in speech are briefly reviewed, and an unsuccessful attempt to replicate the effect using a modified procedure is described.

INTRODUCTION

In both its evolution and its ontogeny, human facility with speech has involved the codevelopment of the abilities to produce and perceive. Since the infant's speech matures to produce the contrasts of his or her native language, some influence whereby perception modifies production is implied. The elucidation of this influence has significance for theoretical accounts of the commonality underlying perception and production (for example, Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967), as well as for the design of programs of speech therapy (for example, McReynolds, Kohn and Williams, 1975).

Empirical support for a link between perception and production can be found in three types of demonstration: first, that perceptual sensitivity to variations in the acoustic properties of speech relates logically to their covariation in production (for example, Summerfield and Haggard, 1977); second, that productive and perceptive capabilities correlate within groups of individuals (for example, Bremer and McGovern, 1977); and third, that an individual's immediate perceptual experience can exert measurable influences on his productions (for example, Lane and Tranel, 1971; Cooper, 1974). The last of these is the most direct and has been shown in several experiments by W. E. Cooper. In these experiments, voice onset times (VOTs) (Lisker and Abramson, 1964) in subjects' productions of voiced and voiceless initial stop consonants were measured after repeated exposure to an adapting sound. The basic result was that VOTs in productions of [pʰi] were shorter following perceptual adaptation with the voiceless adapter [pʰi] than following adapta-

---

[†] Now at the Institute of Hearing Research, The University of Nottingham.

[††] Now at the Department of Psychology, The University of York.

Acknowledgement: This work was carried out by Quentin Summerfield and Peter Bailey while supported by N.A.T.O. post-doctoral research fellowships from the U.K. Science Research Council, and by Donna Erickson while on a post-doctoral research fellowship from N.I.H. We should like to thank Alvin Liberman for his hospitality and encouragement, and Katherine Harris, Seiji Niimi, David Zeichner and Richard Sharkany for advice and assistance during the E.M.C. experiment.

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 2 (1978)]

tion with the isolated vowel [i]. No effect on VOT was found in voiced productions (for example, [bi]) or following exposure to the voiced adapter [bi] (Cooper, 1974). Cooper and Lauritsen (1974) showed that the adaptation effect was not dependent on adapter and test syllable sharing place of articulation: repetitions of [p^{h}i] shortened VOTs in productions of [t^{h}i]. The effect could not be ascribed to mimicry because, overall, subjects produced shorter VOTs following adaptation rather than VOTs approximating that of the adapter. Cooper and Nager (1975) demonstrated that the bisyllabic adapter [rəp^{h}i] could shorten VOTs in productions of both [rəp^{h}i] and [rət^{h}i]. No systematic effects were observed on the duration of either the closure interval or the final stressed vowel, thereby ruling out the possibility that the effects on VOT were due to changes in speech rate or stress. The results of these experiments are summarized in Table 1 where it can be seen that, although they are systematic, the sizes of the changes in VOT are small. The changes correspond to less than 10 percent of the total duration of produced VOTs and are only about half the size of the shifts that can be induced in the phoneme boundary on a VOT continuum by perceptual adaptation [for example, 10.0 msec following adaptation with a voiceless aspirated stop (Eimas and Corbit, 1973)].

Cooper and Nager (1975) concluded that there was a genuine perceptuomotor adaptation effect and that it was to be explained by the fatigue of neural elements that were presumed to be involved both in the perception of voiceless stops and in the abduction of the vocal cords. This "neural model" accounted for the absence of any effect following voiced adaptation and is consistent with the failure of voiceless adapters to change the VOT in voiced productions. The null results of Cooper, Ebert and Cole (1976) are also consistent with the model: none of the adapters [si], [sti], [st^{h}i] and [t^{h}i] systematically affected the VOTs in productions of [sti]. These results were anticipated by the model since the vocal cords are adducted at the moment of stop release in [sti] (Ohala, 1970). Thus, the data in Table 1 show that systematic perceptuo-motor adaptation effects have been obtained for the utterances [p^{h}i], [t^{h}i], [rəp^{h}i] and [rət^{h}i] only after perceptual adaptation with [p^{h}i] or [rəp^{h}i], when compared with the effect of the isolated vowel adapter [i]. One of the objectives of the present study was to determine whether a perceptuo-motor adaptation effect could be obtained with an alveolar adapter for stops produced at all three places of production.

A strong test of the Cooper and Nager (1975) model would require concurrent fiberoptic examination of the vocal cords and electromyographic (EMG) recordings from the intrinsic laryngeal musculature. The model would predict that, following adaptation with a voiceless aspirate stop adapter, the command to abduct the vocal cords would be modified in either or both of two ways. The command either could be weakened so that the cords were abducted to a smaller extent, or could occur later, relative to the adduction command. In both cases, the vocal cords would be less abducted at the moment when adduction commenced, so that the state of approximation required for voicing to onset would be achieved earlier. Preliminary results from a pilot study in which EMG recordings were made from the intrinsic laryngeal musculature are mentioned below, but the cost of such research rendered it necessary to demonstrate that Cooper's basic effect could be obtained reliably and efficiently with a small number of subjects. This was another objective of the present study. Also, since it is desirable to keep testing sessions as short

TABLE 1 : Perceptuo-motor adaptation effects.

Mean reduction in VOT (msec), number of subjects who showed a mean reduction, and significance levels of reductions reported in the sources indicated. Reductions in VOT are indicated by positive values.

	Adapters				
Test Syllable Uttered	[bi]*	[pʰi]*	[rəbi]*	[rəpʰi]*	[sti]+ [stʰi]+ [tʰi]+
[bi][a]	+5.50 9/16 n.s.	+0.30 6/8 n.s.	-	-	-
[pʰi][a]	-0.40 4/8 n.s.	+5.60 13/16 $p < 0.05$	[illegible]	-	-
[di][b]	n.s.	n.s.	-	-	-
[tʰi][b]	n.s.	+3.20 23/32 $p < 0.01$	-	-	-
[rəpʰi][c]	-	-	-	+2.73 12.5/20 $p < 0.05$	-
[rətʰi][c]	-	-	+2.33 13/18 n.s.	+6.50 18/22 $p < 0.001$	-
[sti][d]	-	-	-	-	n.s.

[a] Cooper (1974)
[b] Cooper and Lauritsen (1974)
[c] Cooper and Nager (1975)
[d] Cooper, Ebert and Cole (1976)

* Change in VOT with respect to [i]-adaptation.
+ Change in VOT with respect to [si]-adaptation.

as possible in electrophysiological experiments, we sought in this preliminary study to determine whether the basic perceptuo-motor effect could be obtained with a smaller number of repetitions of the adapter than Cooper had used.

There were three major differences between our procedure and that used by Cooper:

1. We adapted subjects with tokens of their own speech so that each subject heard a different natural speech adapter rather than the same synthetic adapter. We reasoned that any perceptuo-motor link would be most likely to be accessed by a speaker's own speech.

2. We used the bisyllable [rətʰi] as the adapter, rather than [pʰi] (Cooper, 1974; Cooper and Lauritsen, 1974) or [rəpʰi] (Cooper and Nager, 1975).

3. After an initial period of one minute during which 60 repetitions of the adapter were presented, we maintained perceptual adaptation with 10 repetitions prior to each production, following Bailey (1974) and Ganong (1975). Cooper had always presented 70 repetitions of the adapter prior to each production.

This revised procedure did not provide a successful replication of Cooper's basic perceptuo-motor adaptation effect. We shall discuss the extent to which our failure may have resulted from the differences in procedure described above. We present our data to advise others against taking the same course.

METHOD

Subjects

Three female and three male undergraduates served in two experimental sessions lasting about 40 minutes each. They were tested individually and were paid $5.00 for taking part in the experiment.

Procedure

Both sessions of the experiment were conducted in a sound-attenuating booth. Subjects' productions were recorded on magnetic tape with an Ampex AG500 tape recorder and a Shure Model 51 microphone. At the start of the first session each subject practiced uttering the bisyllables [rəpʰi], [rətʰi] and [rəkʰi] in a natural voice with stress on the second syllable. A randomized list was prepared that included 20 instances of each of the three syllables; the list was arranged in columns of 22 items of which the first and last would not be included in subsequent analyses. Each subject was instructed to begin at the top of each column and to utter one token of each item in the list with ten seconds between each item in time with a visual metronome. A brief rest was taken at the end of each column. At the end of the session each subject produced a series of instances of the isolated vowel [i].

We measured two durations in each token. One was the period of devoicing prior to the release of the stop, which we shall refer to as the closure interval; the other was the duration of the VOT. The measurements were made using spectrum and waveform manipulation routines available on the Haskins Laboratories PDP 11/45-GT40 computer system. Preliminary tests showed that the intervals described below were measured most easily when the higher formants had been filtered out. Thus, each token was low-pass filtered at 1.5 kHz, before being digitized with a sampling rate of 10 kHz. The durations were measured from the displayed waveforms by aligning a cursor first with the left-hand boundary and then with the right-hand boundary of the desired interval. The program displays the duration of the demarcated interval. The resolution of the display is variable; with maximum resolution the cursor may be positioned to the nearest 0.2 msec.

Our criteria for measuring the two intervals were as follows: The beginning of the period of devoicing was defined as the peak of the last detectable pitch pulse in the segment [rə] and its end was the beginning of the stop release transient. The VOT was measured from this point to the peak of the first detectable pitch pulse; the first pitch pulse was often difficult to specify. The measures were made on the basis of a consensus of at least two of the three experimenters. From these measures we computed the means and the standard deviations of the closure interval and the VOT for productions of each of the three utterances [rəpʰi], [rətʰi] and [rəkʰi] for each subject. For each subject we then selected that production of [rətʰi] whose VOT was closest to the mean VOT produced by that subject in instances of [rətʰi], and a production of the isolated vowel [i] whose duration was at least as long as the total duration of [rətʰi].[1] These tokens were low-pass filtered at 3.2 kHz and digitized with a sampling rate of 10 kHz. The token of [i] was edited to have the same overall duration as the token of [rətʰi]. A perceptual adaptation tape was constructed from each token in the following format: 60 presentations of the token at a rate of 1/sec were followed by 22 trials each consisting of ten repetitions of the token followed by a three second pause. Thus, there were two tapes for each subject, one each for the adapters [i] and [rətʰi]. These tapes were used in the second session of the experiment.

In the second session each subject listened three times in alternation to each of the tapes constructed from his or her own speech. Subjects rested briefly after each block. The order of presentation of adapters was counterbalanced across subjects. The sequences of adapters were presented binaurally through Grason-Stadler TDH-39 headphones at a constant peak listening level of 80 dB SPL. Subjects were instructed to hold their tongues comfortably against their teeth and not to subvocalize during presentation of the adapters. Immediately after each block of ten adapter repetitions, subjects uttered one of the syllables [rəpʰi], [rətʰi] or [rəkʰi] according to a printed randomization. These productions were recorded as in Session 1. In total, each subject produced twenty tokens of each of the three syllables in both adapter conditions. [Cooper (1974) recorded 20 productions in each condition; Cooper and Nager (1975) recorded 10.] Periods of devoicing and VOTs were measured in the manner described above.

---

[1] The closure durations and VOTs in the [rətʰi] adapters selected for subjects 1 to 6 were:

S1: 102.2, 104.6   S4: 105.2, 72.9
S2: 134.0, 82.3   S5: 95.5, 80.4
S3: 202.3, 121.5   S6: 96.9, 105.9

RESULTS

The results of the experiment are summarized in Tables 2 and 3. Table 2 shows means and standard deviations of VOTs produced by each subject in each syllable for (a) the preadaptation condition in Session 1; (b) after adaptation with [i]; and (c) after adaptation with [rət^{h}i]. Table 3 shows analogous measures for the closure intervals.

The mean VOTs were examined in an analysis of variance with the factors Subjects[6] x Conditions[3] x Test Syllables[3]. The effect of Conditions was not significant ($F_{2,10} = 1.256$; $p < 0.2$) indicating that there was no systematic perceptuo-motor adaptation effect. Overall, and for five of the six subjects, VOTs in [rəp^{h}i] were slightly longer after adaptation with [rət^{h}i] as compared to adaptation with [i]. VOTs in [rək^{h}i] were also slightly longer; three subjects showed this tendency. Only VOTs in [rət^{h}i] were reduced overall by perceptuo-motor adaptation, but only in the productions of three of the subjects. Thus, of the eighteen possible comparisons between effects of [i] and [rət^{h}i] adaptation, eight are in the predicted direction while ten are in the opposite direction. One subject (S3) produced the expected direction of change for all three syllables, while two others (S5 and S6) produced the reversed effect for all three syllables. The Z-scores were computed for each pair of means between the [i] and [rət^{h}i] adaptation conditions. Four of these are sufficiently large to have occurred by chance with a probability less than 0.05; two (for S3 with productions of [rəp^{h}i] and [rək^{h}i]) indicate that a significant reduction in VOT occurred, while two (for S5 with [rək^{h}i] and for S6 with [rəp^{h}i]) indicate that significant lengthening in VOT occurred.

Three other analyses of variance were carried out to examine the differences among standard deviations of the VOTs and among the means and the standard deviations of the closure intervals. A significant effect of Condition was found for the mean closure intervals ($F_{2,10} = 4.15$; $p < 0.05$) indicating that significantly longer closures were produced in the preadaptation condition. This is consistent with a nonsignificant tendency for longer VOTs to have occurred in that session, and suggests that subjects spoke more quickly in the adaptation conditions. The standard deviations of neither the VOTs nor the closure intervals differed systematically between the three conditions.

DISCUSSION AND CONCLUSION

Our data are notable for the complete absence of any differential effect of the two adapters. The mean difference between VOTs produced following [i] and [rət^{h}i] is only 0.27 msec. We shall consider four possible reasons for the difference between this outcome and those reported by Cooper. They are: first, that we did not test a sufficiently large number of subjects; second, that a perceptuo-motor adaptation effect cannot be produced following repeated listening to a subjects' own speech; third, that the perceptuo-motor adaptation effect, though present, failed to emerge because our subjects articulated too carefully; fourth, that we used too few adapter repetitions on each trial. Of these possibilities we consider the last two to be the most reasonable.

TABLE 2 : Voice onset times.

Means and standard deviations of VOTs in milliseconds for each of six subjects in the pre-adaptation condition (Session 1) and the [i]- and [rətʰi]-adaptation conditions (Session 2). Z-scores and their probabilities of occurrence are shown for the comparison between the two means from Session 2.

Subject	Test	Condition						Z-score	P
		Pre-Adap		[i]-Adap		[rətʰi]-Adap			
	Syllable	Mean	S.D.	Mean	S.D.	Mean	S.D.		
1	[rəpʰi]	74.44	10.06	75.51	11.29	75.94	13.75	+0.11	
	[rətʰi]	103.8	7.95	107.4	9.80	103.8	7.38	-1.32	
	[rəkʰi]	104.4	8.36	102.4	11.51	99.54	6.72	-0.96	
2	[rəpʰi]	82.88	12.06	62.71	12.43	62.15	14.53	-0.13	
	[rətʰi]	83.93	8.60	79.81	12.10	75.15	7.91	-1.44	
	[rəkʰi]	113.1	12.77	101.5	8.36	106.9	16.71	+1.29	
3	[rəpʰi]	110.5	19.36	92.59	8.45	85.95	11.81	-2.12	< 0.025
	[rətʰi]	123.0	18.99	110.5	11.38	106.7	8.07	-1.21	
	[rəkʰi]	144.2	15.66	125.7	7.73	119.1	10.43	-2.29	< 0.025
4	[rəpʰi]	57.58	10.67	75.86	10.76	80.65	13.59	+1.24	
	[rətʰi]	72.88	10.07	82.48	5.30	78.86	7.84	-1.71	
	[rəkʰi]	88.33	9.82	101.2	9.42	98.83	10.45	-0.75	
5	[rəpʰi]	56.44	7.32	54.44	10.03	55.44	11.18	+0.30	
	[rətʰi]	80.18	9.41	72.64	6.50	72.71	6.87	+0.03	
	[rəkʰi]	78.75	11.27	72.76	5.71	78.89	10.27	+2.33	< 0.010
6	[rəpʰi]	98.95	8.96	79.91	10.28	86.46	9.41	+2.10	< 0.025
	[rətʰi]	106.2	9.24	95.83	8.54	99.33	11.43	+1.10	
	[rəkʰi]	117.6	8.58	110.3	10.99	112.4	9.72	+0.64	
Means	[rəpʰi]	80.13	11.41	73.50	10.54	74.43	12.27		
	[rətʰi]	95.00	10.71	91.44	8.94	89.42	8.25		
	[rəkʰi]	107.7	11.00	102.3	8.95	102.6	10.72		

TABLE 3 : Periods of devoicing.

Means and standard deviations of periods of devoicing in milliseconds for each of six subjects in the pre-adaptation condition (Session 1) and the [i]- and [rətʰi]-adaptation conditions (Session 2). Z-scores and their probabilities of occurrence are shown for the comparison between the two means from Session 2.

		Condition							
Subject	Test	Pre-Adap		[i]-Adap		[rətʰi]-Adap		Z-score	P
	Syllable	Mean	S.D.	Mean	S.D.	Mean	S.D.		
1	[rəpʰi]	130.0	8.41	93.75	7.95	95.97	10.75	+0.74	
	[rətʰi]	109.7	9.22	74.96	8.92	70.80	10.28	-1.37	
	[rəkʰi]	105.6	5.95	89.30	10.13	89.66	9.98	+0.11	
2	[rəpʰi]	129.9	10.37	100.4	9.84	108.0	12.17	+2.17	< 0.025
	[rətʰi]	136.5	8.39	105.8	15.04	123.0	14.63	+3.67	< 0.001
	[rəkʰi]	103.4	14.42	77.33	19.88	86.59	14.82	+1.67	< 0.050
3	[rəpʰi]	229.5	30.29	158.8	13.13	151.8	17.28	-1.44	
	[rətʰi]	239.3	26.62	166.4	20.36	161.1	14.54	-0.95	
	[rəkʰi]	216.5	31.50	141.4	17.56	127.3	13.60	-2.78	< 0.010
4	[rəpʰi]	94.95	11.09	95.68	13.38	96.70	7.13	+0.30	
	[rətʰi]	93.43	13.01	104.5	10.14	104.2	6.56	-0.11	
	[rəkʰi]	84.20	13.05	87.33	8.75	92.15	7.67	+1.85	< 0.050
5	[rəpʰi]	94.05	12.23	73.51	9.00	79.12	10.56	+1.81	< 0.050
	[rətʰi]	84.88	15.76	62.73	8.50	69.90	11.51	+2.24	< 0.025
	[rəkʰi]	87.03	9.65	65.50	7.53	69.70	9.27	+1.57	
6	[rəpʰi]	125.7	17.71	120.2	12.11	118.6	11.01	-0.44	
	[rətʰi]	130.7	20.69	123.9	13.90	130.6	13.05	+1.57	
	[rəkʰi]	126.8	16.24	118.0	12.00	123.1	14.18	+1.23	
Means	[rəpʰi]	134.0	15.01	107.1	10.90	108.4	11.48		
	[rətʰi]	131.4	15.62	106.4	12.81	109.9	11.76		
	[rəkʰi]	120.6	15.14	96.48	12.64	98.08	11.59		

We do not think that our failure to obtain the effect results solely from our having tested only six subjects. Table 1 reveals that 74 percent of Cooper's subjects showed perceptuo-motor adaptation effects when adapted and tested with voiceless stops. If this figure represents the likelihood of obtaining the expected effect with our modified procedure, then the _a priori_ probability of obtaining it in only one out of six listeners is less than one percent.

We can think of no reason why repetitions of a subjects' own speech should not produce perceptual adaptation. Indeed, in a study of motor-perceptual adaptation (Cooper, Blumstein and Nigro, 1975), the largest perceptual adaptation effect occurred when subjects could hear their own productions. Accordingly, there is no reason to suppose that a subjects' own speech should fail to produce perceptuo-motor adaptation if, as Cooper suggests, the two are linked (Cooper and Nager, 1975; Cooper et al. 1975). However, the absence of perceptuo-motor adaptation would be anticipated by an account which suggests that one consequence of perceptual adaptation is a retuning of motor control such that subsequent productions tend to mimic the coordination of articulatory events that would be entailed in producing the adapter. This explanation would predict no perceptuo-motor consequences of adaptation with a listener's own speech, and, in positing a mimicry of holistic motor coordination, need not predict mimicry of the VOT value of the adapter. However, it is not clear from this account why perceptuo-motor adaptation should be characterized by a particular direction of VOT change.

The overall mean standard deviations of VOTs in [rətʰi] produced after [i] and [rətʰi] adaptation were 8.94 msec and 8.25 msec, respectively, in the present study. They were 9.54 msec and 10.20 msec, respectively, after [i] and [rətʰi] adaptation in Cooper and Nager (1975). Therefore, our failure to find the effect cannot be attributed to our subjects producing more variable VOTs than Cooper and Nager's subjects did. However, our listeners produced both larger closure intervals and longer VOTs than did those of Cooper and Nager, suggesting that they spoke more slowly and that they may have articulated more carefully. The longer period of closure may have allowed time for an albeit weakened "abduction command" to achieve complete vocal cord abduction, so that the relative timing of stop release and the onset of voicing was like that in unadapted productions. While this point would deserve attention in any further study, we note that the single subject in this experiment who consistently showed the expected effect [S3] also produced the longest closure intervals and longest VOTs of the six subjects.

In our attempts to render the perceptuo-motor adaptation procedure more efficient, we may have presented too few repetitions of the adapters. The results of Bailey (1974) suggested that the amount of perceptual adaptation is not systematically increased by increasing the number of adapter repetitions beyond eight. However, in better controlled studies, Hillenbrand (1975) and Simon[2] have shown that although the effect is well established after only ten repetitions, the amount of adaptation does in fact increase with the number of adapter repetitions. It may be that perceptuo-motor adaptation occurs only when perceptual adaptation is maximal, as would have been likely after the 70 adapter repetitions in Cooper's procedure.

---

[2] Simon, Helen. (1977) Anchoring and selective adaptation of phonetic and nonphonetic categories in speech perception. Unpublished Ph.D. dissertation, City University of New York.

As was noted above, the present experiment was undertaken as a precursor to a test of the Cooper and Nager (1975) model by a direct examination of laryngeal behavior. We have carried out a pilot experiment designed to determine whether any changes occur in either the magnitude or the timing of electrical activity of the abductor and adductor muscles in the larynx in productions of voiceless stops following perceptual adaptation. In that experiment we copied Cooper's procedure of having 70 repetitions of a synthetic adapter precede each production. Our adapters were exemplars of the syllables [i] and [rəkʰi]. Two of the experimenters served as subjects and uttered [rəkʰi] after each sequence of adapters. Both subjects showed an effect in the expected direction[3], but technical problems precluded any systematic interpretation of the EMG data. Our experience as subjects emphasized the need to develop a more efficient and thereby more comfortable procedure for use with future subjects.

Clearly, we have not optimized the perceptuo-motor adaptation procedure in the present experiment. In retrospect, while little or no advantage may have derived from adapting subjects with their own productions, our attempts to abbreviate the procedure by reducing the number of adapter repetitions so dramatically may have been counter-productive. We hope that these data will not discourage others from studying perceptuo-motor adaptation effects in speech. Systematic effects have been obtained only with voiceless bilabial adapters (Cooper, 1974; Cooper and Lauritsen, 1974; Cooper and Nager, 1975), and the need to establish the generality of the effect remains.

SUMMARY

Cooper (1974) demonstrated that an effect of perceptuo-motor adaptation could be obtained with speech. In his experiments, VOTs in productions of [pʰ] were found to be shorter after repeated listening to [tʰ] or [pʰ]. Cooper and Nager (1975) accounted for this effect by suggesting that the adapter-fatigued neural elements involved both in the perception of voiceless stops, and in the abduction of the vocal cords. The present experiment was undertaken as a preliminary to an electrophysiological test of this account. We sought to show that perceptuo-motor adaptation could be obtained for productions of stops at all three places of articulation with a testing procedure modelled on Cooper's, but abbreviated to meet the constraints of electrophysiological experimentation. With this modified procedure, we were not successful in replicating Cooper's basic effect. Of six subjects, only one showed a consistent perceptuo-motor adaptation effect of the type described by Cooper, while two others displayed tendencies in the opposite direction.

---

[3] The means of the VOTs in 20 productions of [rəkʰi] following perceptual adaptation with each of [i] and [rəkʰi] were, for subject PJB, 97.2 msec and 90.2 msec, and for subject AQS, 94.7 msec and 93.6 msec.

REFERENCES

Bailey, P. J. (1974) Procedural variables in speech adaptation. *Speech Perception* Series 2, No. 3, 27-34. (Belfast: The Queen's University of Belfast).

Bremer, C. D. and K. McGovern. (1977) Perception of /w,r,l,y/ by normally articulating and misarticulating children. *Journal of the Acoustical Society of America 61*, S64(A).

Cooper, W. E. (1974) Perceptuomotor adaptation to a speech feature. *Perception & Psychophysics 16*, 229-234.

Cooper, W. E. and M. R. Lauritsen. (1974) Feature processing in the perception and production of speech. *Nature 252*, 121-123.

Cooper, W. E. and R. M. Nager. (1975) Perceptuo-motor adaptation to speech: An analysis of bisyllabic utterances and a neural model. *Journal of the Acoustical Society of America 58*, 256-365.

Cooper, W. E., S. E. Blumstein and G. Nigro. (1975) Articulatory effects on speech perception: a preliminary report. *Journal of Phonetics 3*, 87-98.

Cooper, W. E., R. R. Ebert and R. A. Cole. (1976) Speech perception and production of the consonant cluster [st]. *Journal of Experimental Psychology: Human Perception and Performance 2*, 105-114.

Eimas, P. D. and J. D. Corbit. (1973) Selective adaptation of linguistic feature detectors. *Cognitive Psychology 4*, 99-109.

Ganong, W. F. III. (1975) An experiment on "phonetic" adaptation. *Quarterly Progress Report*, no. 116. (Cambridge: Research Laboratory of Electronics, Massachusetts Institute of Technology), 206-210.

Hillenbrand, J. M. (1975) Intensity and repetition effects on selective adaptation to speech. *Research on Speech Perception 2*. (Bloomington Department of Psychology, Indiana University), 56-137.

Lane, H. and B. Tranel. (1971) The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research 14*, 677-709.

Liberman, A. M., F. S. Cooper, D. P. Shankweiler and M. Studdert-Kennedy (1967) Perception of the speech code. *Psychological Review 74*, 431-461.

Lisker, L. and A. S. Abramson. (1964) A cross-language study of voicing in initial stops: acoustical measurements. *Word 20*, 384-422.

McReynolds, L. V., J. Kohn and G. C. Williams. (1975) Articulatory-defective childrens' discrimination of their production errors. *Journal of Speech and Hearing Disorders 40*, 327-338.

Ohala, J. J. (1970) Aspects of the control and production of speech. *UCLA Working Papers in Phonetics 15*.

Summerfield, A. Q. and M. P. Haggard. (1977) On the dissociation of spectra and temporal cues to the voicing distinction in initial stop consonants *Journal of the Acoustical Society of America 62*, 435-448.

Interdependence of Voicing and Place Decisions For Stop Consonants in Initial Position

Bruno H. Repp

ABSTRACT

The present experiments investigate the function relating the voicing boundary for syllable-initial stop consonants to changes in transitional cues for place of articulation, as well as the function relating place boundaries to changes in voice onset time (VOT). Several practiced listeners are studied in detail. The voicing boundary functions are shown to be fairly irregular in shape and to exhibit large individual differences. Certain trends in the functions are best explained by assuming direct effects of categorical place feature decisions on the voicing decision. Other effects must be ascribed to acoustic stimulus properties whose precise nature remains to be determined. Conversely, decisions about the place feature depend on VOT: alveolar responses are more frequent at short VOTs, labial and velar responses at long VOTs. This highly reliable trend appears to be a continuous function of VOT. Thus, the perceptual dependency between the voicing and place features is bidirectional in nature, and while the dependency of voicing decisions on place cues seems to derive from both the phonetic and the auditory level, the dependency of place decisions on voicing cues seems to be purely auditory in nature.

INTRODUCTION

Ever since Lisker and Abramson (1970), Abramson and Lisker, (1973) demonstrated the important role of voice onset time (VOT) in the voiced-voiceless distinction for stop consonants in initial position, the voicing boundary on a synthetic VOT continuum has been an important dependent variable in speech perception research. One of the many factors that affect the precise location of the voicing boundary is place of articulation. Lisker and Abramson (1970) observed that the boundary occurred at successively longer VOTs as place of articulation changed from front to back, that is, from labial to alveolar (apical) to velar. However, their stimuli were constructed to

---

Acknowledgment: This research was supported by NICHD Grant HD01994 to the Haskins Laboratories, and, in its initial stages, by NIH Grant DE-00202 to the University of Connecticut Health Center. I would like to thank Quentin Summerfield for his many excellent comments and his advice on procedural matters. Helpful comments on earlier versions of this paper were also provided by Alvin Liberman, Michael Studdert-Kennedy, and David Isenberg. I am grateful to Georgann Witte for her help in the data analysis. Some of the results were reported at meetings of the Acoustical Society of America (Repp, 1976c, 1977b, 1977c).

[HASKINS LABORATORIES: Status Report on Speech Research SR-53, vol. 2 (1978)]

approximate natural speech as much as possible and thus varied in a number of parameters. Specifically, the release burst and the duration of the formant transitions varied with place of articulation. The duration of the first formant (F_1)--or, rather, its onset frequency at a given VOT, which usually covaries with duration--has been shown to be an important voicing cue in itself (Stevens and Klatt, 1974; Lisker, 1975; Lisker, Liberman, Erickson, and Dechovitz, 1975; Summerfield and Haggard, 1977), and the duration of the second- and third-formant (F_2 and F_3) transitions may have an additional effect under certain conditions (Lisker et al., 1975). Variations in burst spectrum and amplitude are also likely to influence voicing judgments. The striking dependence of the voicing boundary on place of articulation obtained by Lisker and Abramson (1970) must have been partially due to these acoustic parameters that covaried with the place distinction.

Miller (1977) attempted to replicate the Lisker-Abramson results with synthetic stimuli that had identical transition durations and no initial bursts; information about place of articulation was carried only by the starting points and trajectories of the F_2 and F_3 transitions, which are minimal but perfectly sufficient cues to distinguish the three place categories. In these stimuli, the increase in the voicing boundary with place of articulation was much reduced but nevertheless present: the average boundaries occurred at 24.5, 27.8 and 29.3 msec of VOT for labial, alveolar, and velar stops, respectively.[1] Repp (1976a) found a similar difference between the boundaries for labial and alveolar VOT continua of very similar construction. These findings show that the perception of the voicing feature is not completely independent of the place feature, even if the place cues are greatly reduced and simplified.

At which processing level does the dependency of the voicing boundary on place of articulation originate? In theory, it is possible to distinguish three such levels (see Sawusch and Pisoni, 1974):

(1) Auditory analysis. An explanation at this level would require that some direct psychoacoustic interaction occur between the cues for place and voicing. For example, it might be the case that VOT intervals appear subjectively shorter when the F_2 transition is falling (as in most alveolars and velars) than when it is rising (as in labials).

(2) "Cue separation." It may be that the F_2 and F_3 transitions, which primarily carry information about place of articulation, also constitute weak cues to voicing. For example, a high starting frequency of the F_2 transition may bias the voicing decision towards "voiced," without directly affecting the perception of VOT.

(3) Feature decisions. The voicing decision might depend on the outcome of a prior decision about the place category.

---

[1] Positive values of VOT indicate that voicing onset follows the release of the consonant, while negative values indicate prevoicing. Since, in the present context, only positive VOT values are of interest, the customary plus signs are omitted.

The first two levels are difficult to distinguish experimentally and may be considered together as the "auditory" level, while the third level may be called "phonetic" (see Studdert-Kennedy, 1976). Consequently, perceptual interactions at these two hypothetical levels will be called auditory and phonetic, respectively. This corresponds to Smith's (1973) distinction between "stimulus-conditional" and "response-conditional" effects.

Haggard (1970) concluded from an analysis of dichotic confusions that the dependency of the voicing boundary on place of articulation arises at the phonetic level. This implies that place decisions are frequently made before voicing decisions. This was a reasonable assumption for Haggard's stimuli, whose voicing distinctions were carried not by VOT but by a secondary, somewhat artificial cue ("pitch skip") that might have required relatively long processing times. Similarly, the effect of place on voicing decisions reported by Lisker and Abramson (1970) and Miller (1977) was obtained with stimuli that were not ambiguous with respect to place, whereas the voicing feature was naturally uncertain at VOTs in the critical boundary region. Thus, to the degree that uncertainty implies longer decision times, place decisions may be assumed to have occurred before voicing decisions in these studies, too, at least for stimuli in the critical boundary region.

Two recent studies investigated the relationship between voicing and place decisions for stimuli that were ambiguous on both dimensions.[2] Sawusch and Pisoni (1974) tested some simple mathematical models predicting the responses to stimuli on a bidimensional /ba/-/ta/ continuum from the responses to separate unidimensional /ba/-/pa/ and /ba/-/da/ continua. These authors concluded that the perception of the two features was not independent, but the level of the dependency could not be pinpointed in their study. Their data were reanalyzed by Oden (in press) using a more sophisticated mathematical model that assumed complete independence of features. Oden's model seemed to fit the data well, suggesting no perceptual interaction between the place and voicing features. However, the dependence of the voicing boundary on place of articulation was not sufficiently considered in this analysis.

A different approach was taken by Draper (1974); [see also Draper and Haggard, (1974)]. He presented a single syllable over and over, its acoustic parameters having been carefully selected to be ambiguous with respect to both voicing and place. The response distribution showed a correlation between the two features, which provided strong evidence for a perceptual dependency at the phonetic level, since the acoustic structure of the stimulus did not vary. However, Draper noted that this result could reflect a dependency of place decisions on voicing, as well as--or instead of--a dependency of voicing decisions on place. Draper and Haggard (1974) pointed out that although a unidirectional dependency of voicing on place would be in accord with articulation--longer VOTs are a consequence of occlusion in the back of the vocal tract (see Summerfield, 1975b), and certainly not vice versa--there is no reason why perception should not take such an existing articulatory correlation into account and use it also in the reverse direction to infer a posterior place of articulation from relatively long VOTs. Since both features were ambiguous, voicing decisions may have preceded place decisions as well as the reverse, so that a phonetic dependency may have occurred in either direction. Besides yielding equivocal results with respect to the direction of this effect, Draper's paradigm does have some weaknesses: the repeated presentation of a single ambiguous stimulus creates a rather unconstrained situation that is maximally susceptible to response biases, adaptation, and verbal transformation effects (see Goldstein and Lackner, 1974; Ades, 1976), although a more sophisticated tracking procedure employed in Draper's last experiment may have overcome these potential problems. More importantly, however, the design cannot reveal any potential dependencies between voicing and place at the auditory level because the acoustic information is held constant. Thus, the question about the precise nature of the perceptual dependency between voicing and place must remain unresolved.

---

[2] See the General Discussion session for a discussion of the recent work by Oden and Massaro (1977) that was independently conducted at the same time as the later experiments in the present series.

The purpose of the present experiments was to trace in detail the function relating the voicing boundary to the place dimension, as well as the function relating place boundaries to the voicing dimension, by using stimulus arrays varying in both VOT and formant transitions. Two basic hypotheses about the shapes of these functions are illustrated schematically by continuous and broken lines in Figure 1. The more nearly horizontal function represents the voicing boundary (in msec of VOT) plotted as a function of quasi-continuous changes in the primary acoustic cues for place, the F_2 and F_3 transitions. (Only the continuously increasing onset frequency of F_2 is represented by the abscissa; the onset frequency of F_3 first increases and then decreases.) The more nearly vertical functions represent the two place boundaries (labial-alveolar and alveolar-velar, expressed in terms of F_2 onset frequency in Hz) plotted as a function of VOT. Assuming that the voicing boundary function increases from left (labial) to right (velar), a purely auditory hypothesis predicts that it should increase monotonically (the solid function in Figure 1), because the formant transitions constitute a direct cue to voicing.[3] A purely phonetic hypothesis, on the other hand, postulates that voicing decisions depend on place decisions, particularly when there is little uncertainty about the place feature. Therefore, the voicing boundary should not change as long as the place feature remains the same (within place categories), but it should rapidly increase as the place feature value changes from one category to the next (between place categories). Thus, a step function is predicted (the dashed function in Figure 1). In addition, the phonetic hypothesis predicts that the increase in the voicing boundary at place category boundaries would be solely due to the mixture of responses with two different places of articulation; if voicing boundaries were calculated separately for each response place category, they should not differ from the corresponding within-category boundaries, so that the discontinuous place-conditional voicing boundary function should exhibit truly quantal jumps from one place category to the next. Of course, it is possible that both models are correct, that is, that feature interactions take place at both the auditory and the phonetic level; in this case, the voicing boundary function should be monotonically increasing but still exhibit steeper slopes in the region of the place category boundaries. Analogous predictions of the auditory and phonetic hypotheses for the shapes of the place boundary functions are also illustrated in Figure 1. Their direction of change with VOT has been assumed to conform to the suggestion of Draper and Haggard (1974) that listeners use the articulatory correlation between place and VOT in both directions; however, it was not certain whether place boundaries would change at all as a function of VOT.

---

[3] This model includes the possibility that the auditory place information is first transformed into a higher-order code representing the degree to which a stimulus possesses one or the other place feature (see Repp, 1976a; Oden, in press) before it influences the voicing decision.

Underlying the phonetic model are the assumptions that the voicing and place features are processed in parallel, and that the decision time for each feature depends on the nature and relative ambiguity of the acoustic information available. It should be noted that, by assuming sufficiently broad overlapping distributions of decision times for the two features, the phonetic hypothesis can be made nearly indistinguishable from the auditory hypothesis. Thus, a continuously increasing voicing or place boundary function cannot completely rule out a phonetic dependency, while a clear step function would make a purely auditory feature dependency seem unlikely. On the other hand, any irregularities or nonmonotonic trends that might be found in empirical voicing or place boundary functions would have to be attributed to auditory effects, unless they show a clear relation to similar irregularities in the perception of the other feature.

EXPERIMENT I

Method

Subjects. Three volunteers with some experience in listening to synthetic speech were selected because of their consistent performance in earlier experiments. They were paid for their services. The author, a highly experienced listener, also served as a subject.

Stimuli. The stimuli were generated on the OVEIIIc serial resonance synthesizer at Haskins Laboratories. All syllables were 300 msec in duration and had a constant fundamental frequency of 100 Hz. The initial 40 msec contained linear transitions of the three lowest formants from selected starting values to steady states appropriate for the vowel /a/ (F_1 = 77[illegible] Hz, F_2 = 1233 Hz, F_3 = 2520 Hz) which were maintained throughout the rest of the stimulus. (F_4 and F_5 were fixed.) All parameter values were updated every millisecond; together with the fine formant frequency resolution of the synthesizer this led to maximally smooth formant transitions. The starting frequency of F_1 was fixed at 285 Hz for all stimuli. The onset frequencies of F_2 and F_3 covaried in twelve (unequal) steps along a "place continuum," as shown in Table 1. At least nine of these stimuli were expected to be unambiguous with respect to place of articulation; the remaining three (series 4, 5 and 9) fell in the vicinity of place boundaries.[4]

---

[4] In the last two place series, the F_2 and F_3 transitions were extremely close to each other or even crossed at onset. This led to an artifactual click that seemed, however, to have little effect on voicing judgments. This artifact was eliminated in later experiments.

Table 1: Transition onset frequencies in Experiment I.

Series	F_2 onset (Hz)	F_3 onset (Hz)
1	847	1847
2	924	2015
3	993	2181
4	1156	2520
5	1316	2870
6	1467	3199
7	1543	3019
8	1623	2870
9	1770	2520
10	1916	2181
11	2001	2015
12	2075	1847

Table 2: Comparison of place-conditional voicing boundaries (Experiment I).

Subject	F_2 onset (Hz)	Voicing boundary (msec of VOT) Labial	Alveolar	Velar	Predicted
BHR	1156	28.7 (0.51)	30.8 (0.90)	30.1 (0.59)	L A;A V;L V
BHR	1316	-----------	30.5 (0.37)	31.3 (0.69)	A V
BHR	1770	-----------	30.2 (0.58)	31.5 (0.34)	A V
JK	1770	-----------	26.9 (0.72)	29.1 (0.46)	A V
SE	1770	-----------	25.9 (0.60)	28.3 (0.60)	A V
WW	1316	28.3 (0.91)	30.3 (0.86)	-----------	L A

Note: Standard errors in parentheses.

For each of these transition specifications, a series of syllables with different VOTs was synthesized. A delay in voicing onset was produced by substituting low-amplitude hiss for buzz excitation and setting the bandwidth of F_1 to its maximal value (188 Hz). (It is not possible to "cut back" F_1 in a serial synthesizer.) The buzz generator was turned on one pitch period (10 msec) before voicing onset but was kept at a minimal amplitude; this insured a constant amplitude of the first (actually the second) pitch pulse. The pulse generator, which normally is free-running in OVE synthesizers (see Draper, 1974), had been modified so that the first pitch pulse occurred exactly at the moment the generator was turned on.

The VOTs used were: 16, 20, 24, 26, 28, 30, 32, 34, 38 and 42 msec. The four center values (26-32 msec), which spanned the voicing boundaries of all subjects tested, were spaced more closely and presented three times as often as the other six values, leading to a basic set of 18 stimuli per series. Three different randomized sequences of the resulting 12 x 18 = 216 stimuli were recorded. The recordings were made directly from the synthesizer, so that each token of each stimulus represented a new realization of the synthesis parameters. Thus, there were nine different tokens of each of the stimuli in the voicing boundary region and three tokens of each of the other stimuli.

Two analogous stimulus sets were constructed, differing from the original set only in fundamental frequency (F_0). The new F_0s were 80 Hz and 120 Hz, respectively. (The first, inaudible pitch pulse always had a F_0 of 100 Hz, that is, a duration of 10 msec, and occurred 10 msec before the second, audible pitch pulse, so that the latter remained synchronized with voicing onset.) These new stimuli were recorded in the same random sequence as the original set.

Procedure

The tapes were played back on an Ampex AG-500 tape recorder. The subjects listened binaurally over Telephonics TDH-39 earphones at a comfortable intensity. The author listened four times to the original (F_0 = 100 Hz) tape and twice to the other two tapes. The other three subjects only listened twice to the original tape in a single one-hour session.

Analysis

Voicing and place boundary estimates were obtained by fitting psychometric functions to the response percentages using the method of probit analysis (Finney, 1971), and by subsequently taking the 50-percent intercepts of these functions as the boundary estimates. The standard deviations (the reciprocals of the slopes) of these normal ogive functions provided additional parameters of interest.[5] A single place boundary was estimated for the three shortest and the three longest VOTs, respectively; thus, these estimates were based on as many observations as the estimates for each of the four central VOTs.

---

[5] In a number of cases, a complete crossover occurred within only one or two steps on the place continuum. Although the standard deviation of the psychometric function is indeterminate in this case, the computer program used assigned it a certain minimal value greater than 0.

Results and Discussion

Voicing Boundary Functions. The results are shown in Figure 2, separately for the four subjects. Each graph shows the voicing boundary as a function of F_2 starting frequency, as well as the two place boundaries (labial-alveolar and alveolar-velar) as a function of VOT.

First considering the voicing boundaries, it is immediately evident that, while the four subjects did not differ much in their average voicing boundaries, the shapes of the individual voicing boundary functions were extremely different. Moreover, not a single function showed the predicted monotonic (continuous or stepwise) increase with F_2 onset frequency; instead, the functions exhibited peaks and valleys at various unexpected points. Although all subjects had shorter boundaries for alveolars than for velars, they differed in the relationship of the labial boundaries to the boundaries within the other two place categories.

In order to be able to evaluate the two original hypotheses, it must be accepted that the overall shape of the voicing boundary function varies from subject to subject. Once this concession is made, the results of subject SE attract immediate attention, since they are in close agreement with the phonetic hypothesis: the voicing boundary remained nearly constant within place categories but changed abruptly at place boundaries to the value characteristic of the neighboring place category. The voicing boundary function of subject BHR could also be taken to support the phonetic hypothesis, since major changes occurred only in the vicinity of place boundaries. However, there was an unexpected peak in the labial-alveolar boundary region that remains to be explained. The results of subjects JK and WW clearly did not support a simple phonetic hypothesis. Both showed substantial voicing boundary changes within each place category; note especially the pronounced increase within the velar category shown by both subjects. This increase may have been due to the extreme closeness of F_2 and F_3 at the onsets of these stimuli (see footnote 4). However, the resulting artifactual clicks might have been expected to bias perception in favor of voiceless stops, not the opposite.

The voicing boundary function of subject JK showed a peak at the labial-alveolar boundary that was much more pronounced than the corresponding peak in subject BHR's function. In fact, JK's voicing boundary could not be determined for the fifth stimulus series on the place continuum, which explains the interruption in the function (Figure 2). An explanation of this curious increase in voiced percepts is suggested by a peculiar pattern of place confusions exhibited only by BHR and JK: both subjects tended to hear velars in the labial-alveolar boundary region (22 and 43 percent, respectively, in series 5). Since both subjects had longer voicing boundaries for velars than for alveolars, velar intrusions would have increased the voicing boundary estimates if the boundary depended on place decisions. The phonetic hypothesis predicts that the voicing boundaries for place-ambiguous stimuli, when determined separately for each place response category, should equal the respective voicing boundaries for unambiguous stimuli within these place categories, or at least show a difference in the same direction as the difference between the within-category boundaries. Table 2 shows such comparisons for several stimulus series that received a sufficient number of responses in different place categories. Qualitative predictions were derived from the overall trends in the voicing boundary functions of the individual subjects. It can be seen that seven out of eight predictions were confirmed. Moreover, the values of these place-conditional voicing boundaries were generally close to the average values of the adjacent within-category boundaries. Thus, these results support the phonetic hypothesis.

Table 3a provides some additional information about the voicing boundary functions. The average standard deviations and standard errors in the second column show the fairly high accuracy of the listeners in discriminating VOTs in the boundary region. The average uncertainty region (+-2 S.D.) ranged from 10 to 14 msec, and the 99 percent confidence regions around the boundaries (+-2 S.E.) were from 1.2 to 2.4 msec wide. The reliability coefficients in column 3 were obtained by computing the correlations between the results of the two repetitions of the stimulus tape, that is, between the first and second halves of a session.[6] The high coefficients show that the overall trends in the individual voicing boundary functions were highly reliable. Column 4 in Table 3a shows the correlations between the standard deviations and position on the place continuum (F_2 onset frequency). Two subjects showed highly significant negative correlations here, and subject WW did too, if his last two data points, which had abnormally large standard deviations, were excluded. Thus, the psychometric functions tended to get steeper as the place continuum varied from labial to alveolar to velar.

Place Boundary Functions. Figure 2 shows clearly that the place boundaries were not independent of VOT: the two boundaries converged--that is, alveolar responses decreased in frequency--as VOT increased.[7] This trend was shown by all four subjects. The effect was larger than the figure suggests; allowance must be made for the relatively compressed abscissa scale. There was no indication whatsoever that the shifts in the place boundary functions took place only, or primarily, across the voicing boundary. On the contrary, the largest shifts often occurred at short VOTs that were all within the voiced category. Thus, the dependence of place decisions on VOT was apparently not phonetic in nature.

---

[6] In the case of BHR, the correlation was between two whole sessions. The values correlated were the total numbers of voiced responses for each stimulus series along the place continuum in each (half-)session. These values were very closely related (by a linear transformation, with small error) to the boundary estimates obtained from probit analysis.

[7] The data points at VOT = 24 msec and VOT = 34 msec combine the responses at all shorter and longer VOTs, respectively. The actual mean VOTs were 20 and 38 msec, respectively. Thus, all six data points of each place boundary function were based on an equal number of observations. Both place boundaries were derived from the percentages of alveolar responses, so that velar intrusions at the labial-alveolar boundary were grouped with labial responses.

Table 3: Some indices of variation and covariation (Experiment I).

(a) Voicing boundary functions (msec of VOT).

Subject	S.D. (S.E.)	$r_{I,II}$	$r_{S.D.,F_2}$
BHR[a]	2.35 (0.30)	+0.93***	-0.80***
JK	2.99 (0.52)	+0.89***	-0.19
SE	2.45 (0.54)	+0.90***	-0.75**
WW	3.55 (0.59)	+0.88***	+0.12 (-0.73**)[b]

(b) Place boundary functions (Hz of F_2 onset).

Subject	S.D. (S.E.)		$r_{S.D.,VOT}$	
	L/A	A/V	L/A	A/V
BHR[a]	96 (15)	58 (11)	+0.74*	-0.05
JK	78 (20)	67 (18)	+0.98***	-0.64
SE	63 (18)	46 (15)	+0.02	+0.62
WW	66 (19)	49 (16)	+0.90**	+0.21

$^{*}p < .05$
$^{**}p < .01$
$^{***}p < .001$
[a] Data for BHR from two sessions.
[b] Last two data points omitted.

Legend: S.D. = average standard deviation
S.E. = average standard error
$r_{I,II}$ = reliability coefficient (half-session correlation)
$r_{S.D.,F_2}$ = correlation between voicing boundary S.D. and F_2 onset frequency (n = 12)
L/A = labial-alveolar boundary
A/V = alveolar-velar boundary
$r_{S.D.,VOT}$ = correlation between place boundary S.D. and VOT (n = 6)

Table 3b shows some additional statistics about the place boundaries. It can be seen that all four subjects exhibited smaller standard deviations and standard errors for the alveolar-velar (A/V) boundary than for the labial-alveolar (L/A) boundary, even those subjects (SE and WW) who gave no velar intrusions at the L/A boundary. Table 3b also shows the correlations between the standard deviations of the place boundaries and VOT. Since each coefficient was based on only six pairs of observations, there was large variability, but there were three significant positive correlations, indicating a tendency for uncertainty about place to increase with VOT.

Different Fundamental Frequencies. A further opportunity to assess the reliability of individual results was provided by the two stimulus series with different fundamental frequencies. The F_0 parameter was not expected to affect the shape of the voicing boundary function, but an overall shift in the function was considered likely, such that a higher F_0 would lead to more voiceless responses (Massaro and Cohen, 1976). The results for subject BHR are shown in Figure 2a as the dashed functions. As expected, the voicing boundary functions for all three conditions were very similar in shape but at different levels of VOT: stimuli with F_0 = 120 Hz had their boundaries at shorter VOTs, and stimuli with F_0 = 80 Hz had their boundaries at longer VOTs than stimuli with F_0 = 100 Hz. The similarity of the three functions is evident in their high intercorrelations (between +0.84 and +0.88, all $p < .01$). The new voicing boundary functions were also highly reliable (half-session correlations of +0.93 for both) and exhibited considerably smaller standard deviations (average: 1.62 msec) than the F_0 = 100 Hz condition--probably a consequence of practice. The place boundaries (not shown in Figure 2a) basically unaffected by changes in F_0, again showed the convergence as VOT increased, and also had reduced standard deviations (average: 51 Hz).[8]

---

[8] While the overall shapes of the three voicing boundary functions for subject BHR were very similar, there seemed to be little agreement strictly within place categories. This was confirmed in an analysis of variance of the nine within-category boundaries (series 1-2-3; 6-7-8; 10-11-12 on the place continuum), with the factors: F_0 (3 levels), place categories (3), within-category steps (3) and (half-)session replications (2). The pooled interactions of the replications factor with all other factors were taken as the error estimate. The effects of F_0 and of place categories were highly significant ($F_{2,26} = 186.3$, $p << .01$, and $F_{2,26} = 338.1$, $p << .01$, respectively), but the within-category effect did not reach significance, nor did any of its interactions with other factors. This lent further support to the phonetic hypothesis. The statistical analysis also revealed a significant interaction of F_0 and place categories ($F_{4,26} = 10.5$, $p < .01$), due to the overlap of the functions for the two lower F_0 frequencies at the velar end of the continuum, and a significant effect of replications ($F_{4,26} = 24.5$, $p < .01$), due to a decrease in voiced responses in the course of a session--an effect also reported by Summerfield (1975a) and consistent with an asymmetry often found in selective adaptation and discrimination of VOT (for example, Eimas and Corbit, 1973).

Summary

The changes in the voicing boundary with place of articulation give moderate support to the phonetic hypothesis. The data of two subjects conformed to the hypothesis; those of the other two did not. The latter two subjects showed apparently reliable voicing boundary changes within place categories; however, these changes did not follow the smooth monotonic course predicted by the auditory hypothesis outlined in the Introduction. While the effects must have been auditory in nature, they were clearly not a direct function of F_2 onset frequency.

All subjects showed a decrease in alveolar responses as VOT increased. This decrease clearly did not depend on the voicing boundary and appeared sufficiently regular for an auditory explanation to apply.

EXPERIMENT II

Experiment II had several purposes. First, it attempted to gain even more precise information about the shapes of voicing and place boundary functions. In order to achieve this, the stimulus series of Experiment I was extended from 12 to 24 steps along the place continuum. The velar end of the place continuum was shortened, in order to avoid artifacts due to close proximity of F_2 and F_3 onsets (see footnote 4). Instead of presenting the same stimulus tape twice, as in Experiment I, two tapes were recorded, so that each individual stimulus was a different token from the sampling distribution generated by the synthesizer.

A second purpose of the experiment concerned the peaks at the labial-alveolar boundary for subjects BHR and JK. Although the coincidence of these peaks with velar intrusions was very convincing, there could be an alternative explanation. Since the steady-state F_2 and F_3 frequencies of the /a/ vowel were 1233 Hz and 2520 Hz respectively, it so happened that both F_2 and F_3 were nearly flat in the stimulus series where the peak occurred (see Table 1). It could be that some subjects were sensitive to the absence of transitions in the higher formants and succumbed to some psychoacoustic effect that biased them towards hearing the stimuli as more voiced. In order to provide a test of this hypothesis, three additional stimulus series were constructed by varying the steady-state of F_2. The F_2 onset frequencies and F_3 were left unchanged. The F_2 steady states were chosen so as to lead to flat transitions in different regions of the place continuum. If it was flatness of F_2 that caused the peak in the voicing boundary function, then the original peak should disappear and a new peak should be found wherever F_2 happened to be flat. If the peak was caused by velar intrusions, the original peak should remain, since varying the F_2 steady-state was not expected to affect the probability and location of velar intrusions; and if it did, the peak should shift with the intrusions. (It was assumed that F_2 was the important factor. Since F_3 remained unchanged, a fixed peak in the region of the labial-alveolar boundary could conceivably be due to flatness of F_3 only, but this possibility was considered rather remote.)

Experiment II provided a large amount of data that could be used to assess the reliability of local trends in the voicing boundary functions. The replicability of such local variations across conditions with different F_2 steady-state frequencies was of special interest. The procedure of varying F_2 was also expected to affect the location of the place boundaries.[9] This provided an elegant test of the phonetic hypothesis: any place-conditional changes in the voicing boundaries should shift with the place boundaries, if these changes are truly phonetic in nature.

Method

Subjects. Three of the four subjects of Experiment I participated in this experiment: BHR, JK and SE.

Stimuli. A 24-step place continuum was generated on the OVEIIIc synthesizer. Except for the more closely spaced F_2 and F_3 onset frequencies, the new stimuli were identical to those in Experiment I (F_0 = 100 Hz). The new transition onset frequencies are shown in Table 4. The 24 stimulus series fell into four groups of six, as indicated in Table 4. Groups 1 (series 1-6) and 3 (series 13-18) were "within-category" groups, intended to be perceived as all-labial and all-alveolar, respectively. Groups 2 (series 7-12) and 4 (series 19-24) were "between-category" groups, since they were expected to span the labial-alveolar and alveolar-velar boundaries, respectively. Step sizes varied somewhat from group to group. Note that, in series 24, the onsets of F_2 and F_3 were still separated by almost 300 Hz, so that no artifactual clicks occurred (see footnote 4).

The VOTs were the same as in Experiment I. Four experimental tapes were recorded. Each of two "within-category" tapes contained three blocks of 216 stimuli, each consisting of a different randomization of the combined stimuli from the two within-category groups. Each of the other two ("between-category") tapes contained three similar blocks of stimuli from the between-category groups. The ISI interval was reduced to 2 sec.

Three additional stimulus series were created, identical to the one just described, except that the F_2 steady-state was changed to either 924 Hz (very low), 1565 Hz (medium), or 1796 Hz (high). (The original F_2 frequency, 1233 Hz, was considered low.) The three new F_2 frequencies led to changes in vowel quality: the new vowels were approximately /ʌ/, /ae/, and a very open /ae/, respectively. Transition duration remained constant at 40 msec. Thus, although the starting frequencies of F_2 were held constant (Table 4), the whole time course of the F_2 transitions varied with the steady-state of F_2. For each of the three new stimulus sets, four experimental tapes were recorded

---

[9] Such shifts are predicted by the theory of loci (Delattre, Liberman, and Cooper, 1955) which assumes that the perceptually relevant F_2 onset frequency (the locus) must be extrapolated from the actual onset frequencies backward in time. Consequently, holding F_2 onset constant and raising the F_2 steady-state should lead to an undershoot of the locus, while lowering the F_2 steady-state should lead to overshoot. Thus, as F_2 steady-state increases, both place boundaries should shift to the right on the place continuum, if the theory of loci is correct.

Table 4: Transition onset frequencies in Experiment II.

Group	Series	F_2 onset (Hz)	F_3 onset (Hz)
1	1	829	1808
	2	866	1888
	3	904	1972
	4	944	2059
	5	986	2150
	6	1029	2229
2	7	1067	2295
	8	1139	2413
	9	1207	2539
	10	1279	2651
	11	1345	2789
	12	1415	2891
3	13	1467	2998
	14	1510	3086
	15	1554	3176
	16	1588	3176
	17	1623	3086
	18	1658	2998
4	19	1695	2891
	20	1744	2789
	21	1783	2651
	22	1822	2539
	23	1862	2413
	24	1902	2295

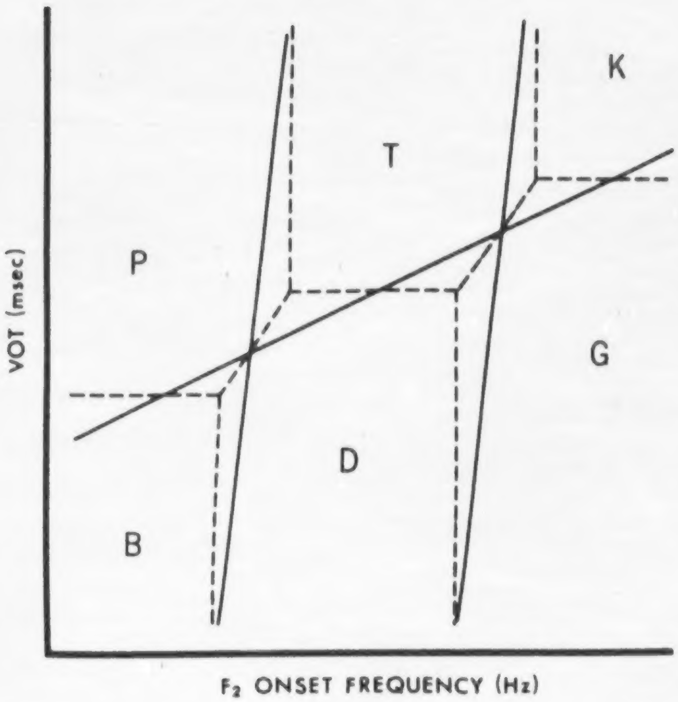


Figure 1: Schematic illustration of two simple hypotheses about the dependency of voicing decisions on transitional cues for place of articulation and of place decisions on VOT. The solid lines represent the auditory model, while the dashed lines represent the phonetic model. The horizontal lines are voicing boundary functions, and the vertical lines are place boundary functions (labial-alveolar and alveolar-velar, respectively). Large letters indicate the areas in which the corresponding stop consonants are the predominant response.

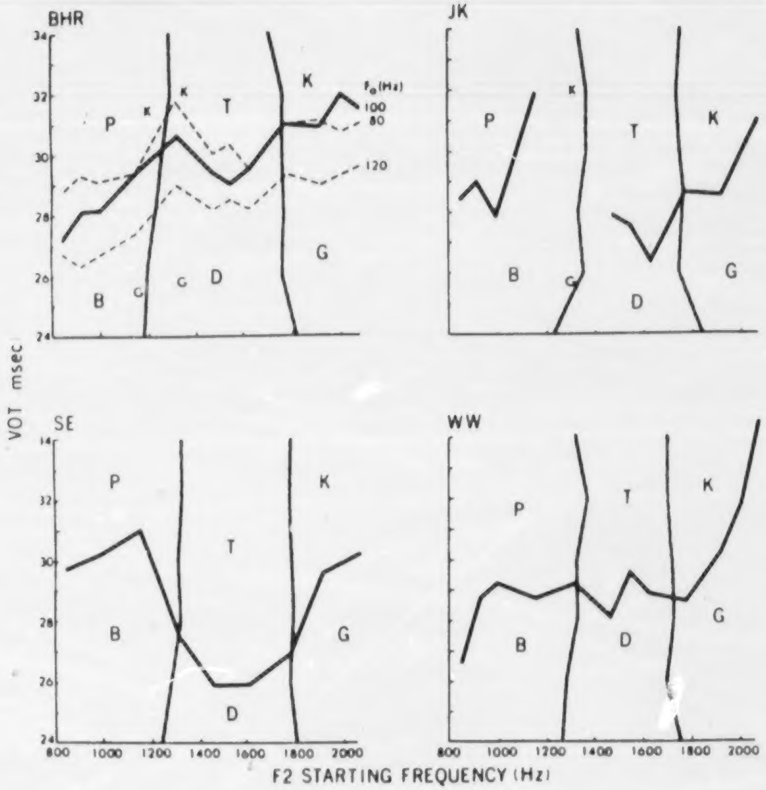


Figure 2: Results of four subjects in Experiment I. Voicing boundaries (solid functions) are shown as a function of changes in F_2 onset frequency (abscissa); place boundaries (dashed vertical functions) are shown as a function of VOT (ordinate). Additional voicing boundary functions for two conditions with different F_0s are shown for subject BHR (dashed horizontal functions, upper left-hand panel). Small Gs and Ks in the labial-alveolar boundary region indicate velar intrusions (subjects BHR and JK).

that were identical with those for the low F_2 series, including the randomization.

Procedure

Each condition required two one-hour listening sessions. In the first session, a within-category tape (three blocks of 216 stimuli) was followed by a between-category tape; in the second session, the other between-category tape was followed by the other within-category tape. The subjects were aware that, on a within-category tape, only labial and alveolar stimuli were expected to occur, but they were encouraged to write down velar responses, if that was what they heard. All subjects served in the low F_2 condition first. Subject BHR repeated both sessions, so that his results were based on twice as many observations as those of the other two subjects. Subsequently, subjects JK and SE listened to the three other conditions in the order: very low, medium, high; subject BHR was assigned the reverse order.

Results and Discussion

Voicing Boundary Functions. The results of Experiment II are shown in Figure 3. The voicing boundary functions for the four conditions with different F_2 steady states have been vertically displaced, in order to facilitate comparisons. The results of the low F_2 condition (second function from bottom), which was basically a replication of Experiment I, showed some disagreement with the earlier results (particularly for subjects JK and SE) due to reduction of extreme peaks and valleys--perhaps a consequence of practice. However, each individual retained his characteristic shape of the voicing boundary function, and subjects BHR and JK again showed pronounced peaks in the labial-alveolar boundary region, together with velar intrusions.

One hypothesis under test predicted that peaks in the voicing boundary function should appear at the point along the place continuum where the F_2 transition was flat. These points are indicated by the arrows in Figure 3. Apart from the expected peak in the low F_2 condition, subject BHR showed a small peak at the predicted location in only one of the three other conditions (very low F_2). Subject JK showed minor peaks in all three other conditions, but his functions were so jagged that this could easily have occurred by chance. Subject SE, who had not shown a major peak in Experiment I, showed a pronounced peak at the predicted location (in the labial region) in the very low F_2 condition, but not in the other two conditions. However, SE's peak in the labial category was not unique to the very low F_2 condition, but appeared in the other conditions as well, although in less pronounced form. Thus, there was no convincing support for peak shifts with changes in F_2 steady state.

On the other hand, it was predicted that the peak near the labial-alveolar boundary obtained in the low F_2 condition would disappear in the other conditions and prove to be unrelated to velar intrusions. This was clearly not the case. Subject JK, in particular, showed pronounced peaks in each condition that exactly coincided with substantial frequencies of velar intrusions. Subject BHR showed a clear peak in the very low F_2 condition which coincided with an unusually high proportion of velar responses (73

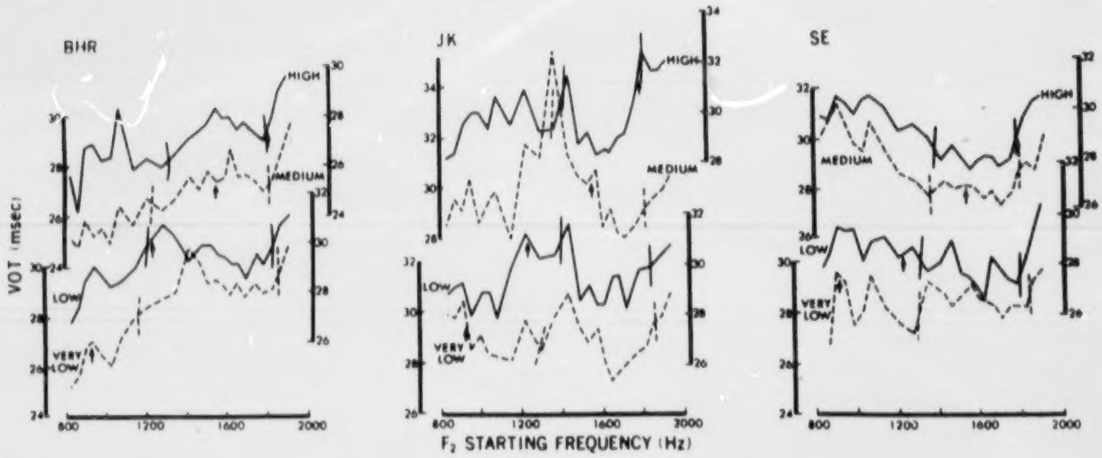


Figure 3: Results of three subjects in Experiment II. Each panel shows, vertically displaced, the voicing boundary functions for the four conditions with different F_2 steady states (from top to bottom: high, medium, low, very low). Of the place boundaries, only the intersections with the voicing boundaries are indicated. (See Tables 5 and 7 for additional information.) Arrows indicate the positions on the place continuum where F_2 was flat.

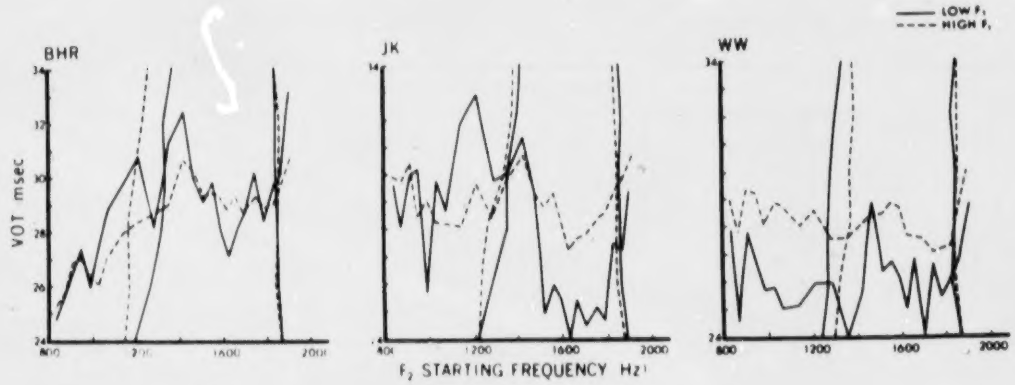


Figure 4: Comparison of low F_1 results (Experiment III) with high F_1 (= very low F_2) results (Experiment II) for three subjects.

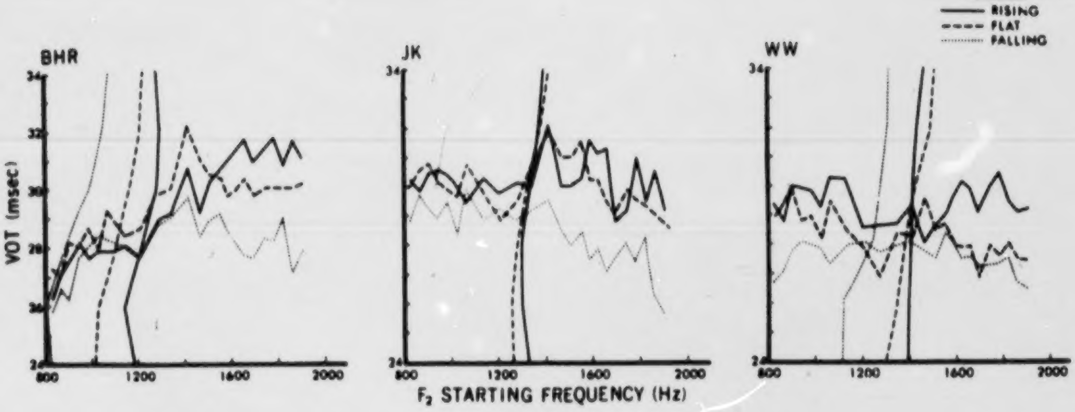


Figure 5: Comparison of three conditions with different fixed levels of F_3 (Experiment IV). For rising F_3, the (single) place boundary separates labial and velar responses (with alveolar intrusions, except for subject WW); for flat or falling F_3, the place boundary separates labial and alveolar responses.

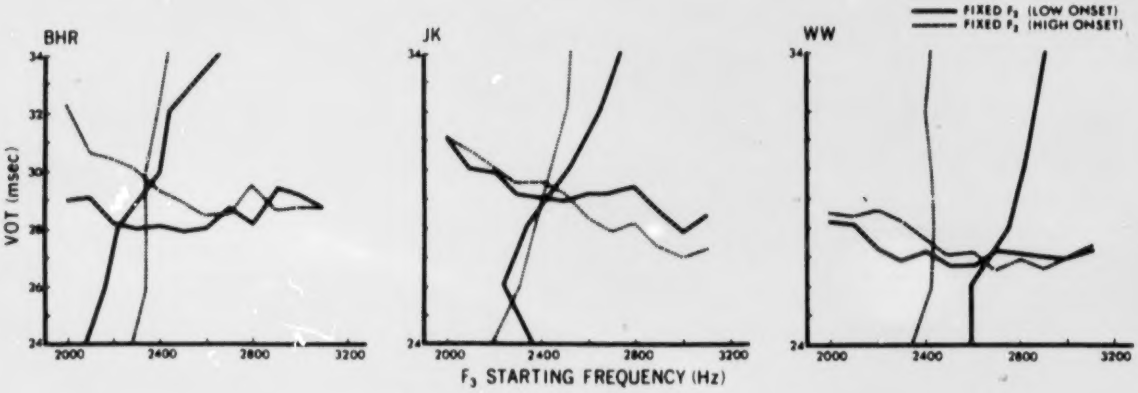


Figure 6: Comparison of two conditions with different fixed levels of F_2, with F_3 onset varying (Experiment V). For low F_2 onset, the place boundary separates labial and alveolar responses; for high F_2 onset, it separates velar and alveolar responses.

percent). In the other two conditions, there were no pronounced peaks, but velar intrusions were also rather infrequent. Subject SE showed no clear peaks, but nor did he give any velar intrusions (except for a few in the very low F_2 condition). Thus, in general, the hypothesis that velar intrusions shift the voicing boundary towards longer VOTs is much better supported by the data than the hypothesis that relates peaks to flatness of F_2.

Table 5a shows some statistics about the voicing boundary functions. The first subtable lists the average voicing boundaries, since level differences in the voicing boundaries are difficult to see in Figure 3 because of the vertical displacement. Only subject BHR showed a substantial effect of F_2 steady-state on voicing boundaries, the boundaries being at longer VOTs for the two lower F_2 steady states. Average standard deviations and standard errors were considerably smaller than in Experiment I, probably due to practice. Average uncertainty regions (+-2 S.D.) around the boundaries were as small as 6-11 msec. Two subjects were most accurate in the very low F_2 condition, but otherwise there was little relation to F_2 steady-state. Between-session reliabilities, shown in the next subtable, were somewhat lower than the within-session reliabilities in Experiment I.[10] In part, this may have been a consequence of random variation introduced by the increase in data points. Reliability did not vary systematically with F_2 steady-state. The correlations between standard deviations and F_2 onset frequency, shown in the final subtable of Table 5a, were quite consistently negative, four of them reaching significance. Thus, as in Experiment I, standard deviations tended to decrease as F_2 onset frequency increased.

It was of particular importance to find out whether variations in the voicing boundary function strictly within place categories were reliable, and if so, whether they were as reliable as changes across place category boundaries. To this end, between-session correlations were obtained separately for the six data points within each stimulus group (see Table 4), leading to two within- and two between-category reliability coefficients per condition per subject--48 coefficients altogether. Of the 24 within-category correlations, 23 were positive ($p < .001$) and 12 were significant at the .05 level. All 24 between-category correlations were positive, 16 significantly so. The average group coefficients (averaged over the four conditions with different F_2 steady-states) are shown in Table 6a. The average within-category reliabilities (groups 1 and 3) were only slightly lower than the average between-

---

[10] In order to take the counterbalancing of conditions into account, the within-category results of the first session were combined with the between-category results of the second session and correlated with the two remaining series. For subject BHR, whose boundaries consistently shifted towards shorter VOTs within a session, this procedure resulted in higher reliabilities than the straight between-session correlations; for the other two subjects, whose boundaries were more stable within than between sessions, the latter correlations were usually somewhat higher. For example, the very low reliability of +0.35 for JK's low F_2 function was due to a large between-session boundary shift for this subject. The straight between-session reliability of the same function was +0.59 ($p < .01$).

Table 5: Some indices of variation and covariation: Voicing boundary functions (Experiment II).

Subjects	F_2 steady state			
	Very low	Low	Medium	High
(a) Average voicing boundaries (msec of VOT).				
BHR[a]	28.40	29.25	26.82	27.12
JK	29.01	29.44	29.85	29.69
SE	28.39	28.24	28.81	29.02
(b) S.D. (S.E.) (msec)				
BHR[a]	1.60 (0.34)	1.94 (0.26)	1.93 (0.41)	1.90 (0.40)
JK	1.77 (0.34)	2.73 (0.46)	2.77 (0.49)	2.27 (0.42)
SE	1.63 (0.33)	1.68 (0.33)	1.63 (0.32)	1.35 (0.29)
(c) $r_{I,II}$ (n = 24)				
BHR[a]	+0.90***	+0.84***	+0.86***	+0.74***
JK	+0.52**	+0.35[b]	+0.61***	+0.61***
SE	+0.65***	+0.63***	+0.95***	+0.79***
(d) $r_{S.D.,F_2}$ (n = 24)				
BHR[a]	-0.73***	-0.73***	-0.02	-0.33
JK	-0.33	-0.18	-0.32	+0.06
SE	-0.14	-0.48*	-0.29	-0.42*

*p < .05
**p < .01
***p < .001
[a] Low F_2 data for BHR based on 4 sessions.
[b] See Footnote 10.

Table 6: Average within-group reliabilities and intercorrelations of voicing boundary functions (Experiment II).

Subject	Stimulus Group				
	1	2	3	4	Whole Functions
(a) Average within-group reliabilities.					
BHR	+0.70	+0.68	+0.40	+0.73	+0.84
JK	+0.70	+0.73	+0.59	+0.73	+0.52
SE	+0.69	+0.73	+0.57	+0.88	+0.76
(b) Average within-group intercorrelations between F_2 conditions.					
BHR	+0.64	+0.09	-0.20	+0.82	+0.74
JK	-0.22	+0.39	-0.08	+0.81	+0.38
SE	+0.44	+0.41	+0.16	+0.81	+0.45

category reliabilities (groups 2 and 4)--a nonsignificant difference. These results establish conclusively that there were significant variations in the voicing boundary within place categories.

The cause of these--often very irregular--variations in the voicing boundary within place categories is not clear. Their interpretation would be greatly facilitated if they also proved reliable across conditions. The answer to this question is provided in Table 6b, which shows, separately for each stimulus group and each subject, the arithmetic average of the six intercorrelations among the four conditions with different F_2 steady states. It is readily apparent that only the sudden increase in the voicing boundary across the alveolar-velar boundary (group 4) was consistent across conditions and subjects--all 12 functions in Figure 3 exhibit this trend. The between-condition correlations of other sections of the voicing boundary function ranged from moderate to zero. Thus, neither the within-category variations in the voicing boundary, nor even the changes across the labial-alveolar boundary were the same in different conditions, although most of these changes were highly reliable within conditions.

Place Boundary Functions. The place boundary functions (only their intersections with the voicing boundary functions are indicated in Figure 3) confirmed the dependency of the place boundaries on VOT as already observed in Experiment I. All subjects in all conditions showed a decrease in alveolar responses as VOT increased, resulting in a marked convergence of the two place boundaries. The labial-alveolar boundary seemed to shift somewhat more with VOT than the alveolar-velar boundary, but both boundaries were clearly affected. Again, there was no evidence that the boundary shifts took place only, or primarily, across the voicing boundary, thus not supporting a phonetic explanation of the effect.

As expected, the locations of the place boundaries were affected by varying the F_2 steady-state frequency. The average boundary locations are shown in Table 7a. It can be seen that, as F_2 steady-state increased, the labial-alveolar boundaries of all three subjects shifted towards higher F_2 onset frequencies, while the alveolar-velar boundaries shifted towards lower F_2 onset frequencies. The former shift was about twice as large as the latter, and (for subject BHR, at least) tended to accelerate, while the latter clearly decreased with increasing F_2 steady-state frequency. Only the shift of the labial-alveolar boundary was in agreement with the predictions derived from the "locus theory" (Delattre, Liberman and Cooper, 1955; see footnote 9), while the shift of the alveolar-velar boundary was in the opposite direction.

The phonetic hypothesis predicted that the steep increase in the voicing boundary at the alveolar-velar boundary would shift, together with the place boundary as F_2 steady-state was changed. The F_2 onset frequencies at which major increases in the voicing boundary occurred (F_L) are also shown in Table 7a. As can be seen, there was only weak support for the prediction. In view of the relatively small shifts of the place boundary, this was perhaps not too surprising. The fact that the major increase in the voicing boundary generally coincided with the alveolar-velar boundary was itself in accord with the phonetic hypothesis. In several instances, however, the increase seemed to continue beyond the place boundary region, as had already been observed in

Table 7: Some indices of variation and covariation:
Place boundary functions (Experiment II).

		F_2 steady state			
Subject	Boundary	Very Low	Low	Medium	High
(a) Average place boundary locations (in Hz of F_2 onset) and F_L values.[a]					
BHR	L/A	1196	1217	1255	1328
	A/V	1857	1822	1807	1804
	F_L	1822	1783	1822	1783
JK	L/A	1296	1359	1315	1373
	A/V	1841	1819	1798	1801
	F_L	1695	1695	1695	1695
SE	L/A	1309	1317	1356	1374
	A/V	1839	1813	1804	1793
	F_L	1822	1822	1744	1744
(b) S.D. (S.E.) (Hz)					
BHR	L/A	66 (12)	47 (7)	63 (12)	79 (13)
	A/V	19 (6)	31 (4)	40 (8)	30 (6)
JK	L/A	101 (16)	63 (11)	85 (14)	88 (15)
	A/V	36 (7)	40 (7)	50 (8)	38 (7)
SE	L/A	63 (12)	40 (9)	44 (10)	42 (9)
	A/V	31 (6)	42 (5)	48 (8)	33 (6)
(c) $r_{S.D.,VOT}$ (n = 6)					
BHR	L/A	+0.50	+0.93***	-0.05	-0.31
	A/V	-0.23	+0.84**	-0.49	+0.59
JK	L/A	+0.25	+0.50	+0.03	-0.18
	A/V	+0.69	+0.78*	+0.45	+0.57
SE	L/A	+0.76*	+0.53	+0.47	+0.31
	A/V	+0.64	+0.62	+0.07	-0.59

*$p < .05$
**$p < .01$
***$p < .001$

[a] F_L = F_2 onset frequency at which the first substantial increase in the voicing boundary occurred (A/V boundary region).

Experiment I. Thus, an auditory explanation in terms of the rapid convergence of F_2 and F_3 cannot be ruled out.

Tables 7b and 7c show some other statistics. The standard deviations and standard errors (Table 7b) were again markedly longer for the labial-alveolar boundaries than for the alveolar-velar boundaries, due in part to velar intrusions. The correlations between the standard deviations and VOT (Table 7c) were predominantly positive. This tendency for place decisions to become less accurate at longer VOTs was also in agreement with Experiment I.

Summary

The results of Experiment II suggest that the dependency of the voicing boundary on place of articulation is both phonetic and auditory in nature. The most convincing evidence for phonetic effects comes from the increases in the voicing boundary due to velar intrusions in the labial-alveolar boundary region. To a lesser extent, the rapid increase in the voicing boundary across the alveolar-velar boundary suggests a phonetic effect. No systematic phonetic effects were observed across the labial-alveolar boundary.

The auditory effects observed were not as regular as envisioned at the outset (Figure 1). Rather, they were generally nonmonotonic in nature and consisted of multiple peaks and valleys in the voicing boundary function within the labial and alveolar categories. These irregularities proved to be reliable within conditions, but differed between conditions and between subjects. Apparently, they were due to some very specific aspects of the acoustic structure of the stimuli, possibly reflecting certain characteristics of the synthesizer used.

The dependency of the place boundaries on VOT poses fewer interpretative problems. The convergence of the boundaries as VOT increased was highly consistent across all conditions and subjects, and generally monotonic in character. The absence of any relation to the voicing boundary suggests an auditory explanation.

EXPERIMENTS III-V

In Experiments I and II, the onset frequencies of F_2 and F_3 always varied simultaneously. The resulting change in the relationship between F_2 and F_3 may have been responsible for some of the more complex effects observed. The following experiments looked at the effects of varying a single formant. First of all, Experiment III examined whether changing the F_1 transition affects the shape of the voicing boundary function, or whether the voicing boundary function depends solely on F_2 and F_3 onsets. Subsequently, Experiment IV varied F_2 onset, holding F_3 onset constant at one of three values. Finally, Experiment V varied F_3 onset, holding F_2 constant at one of two values. F_1 remained unchanged in Experiments IV and V. It was hoped that these experiments would lead to more systematic and perhaps more readily interpretable voicing boundary functions than the previous studies, and that some information would be obtained about the source of the variability within place categories.

Method

Subjects. Subjects BHR and JK continued to serve as subjects. Subject SE, who was no longer available, was replaced by WW, who had served as a subject in Experiment I.[11]

Stimuli. New stimulus sets were created by modifying the very low F_2 set of Experiment II (F_2 steady-state of 924 Hz), which had yielded especially accurate performance. For Experiment III, the steady-state of F_1 was lowered from 771 Hz to 500 Hz, resulting in a vowel color close to /ɔ/. F_1 onset remained at 285 Hz, and the onsets of F_2 and F_3 covaried as previously (Table 4).

For Experiment IV, three new stimulus sets were constructed, identical with the very low F_2 sets of Experiment II, except that the onset frequency of F_3 was fixed within each series and only F_2 onset varied. The three onset frequencies of F_3 were 1808 Hz (rising), 2520 Hz (flat) and 3176 Hz (falling). Since a rising F_3 transition is appropriate for labials and velars, while a falling F_3 transition is characteristic of alveolars, pronounced shifts in the place boundaries were expected (see Harris, Hoffman, Liberman, Delattre, and Cooper, 1958; Hoffman, 1958). These changes again provided an opportunity to detect phonetic effects on the voicing boundary.

For Experiment V, two new stimulus sets were created by holding the F_2 transition constant at one of two values (low F_2 onset: 1207 Hz; high F_2 onset: 1822 Hz) and varying F_3 onset from 2000 to 3100 Hz in steps of approximately 100 Hz. The F_2 onsets were chosen to fall in the place boundary regions of the very low F_2 set of Experiment II, so that the F_3 transition would be critical for place distinctions. The two stimulus sets were treated as if they were the within- and between-category groups of a single 24-step place continuum.

Procedure

All stimulus sets were recorded and presented exactly as in Experiment II. Subjects JK and WW did the experiments in the order IV-V-III; subject BHR did them in the order IV-III-V.

Results and Discussion

Experiment III: Lowering F_1 Steady-state. The results for the low F_1 stimulus series are shown in Figure 4 (solid functions), together with the results of the very low F_2 (= high F_1) condition from Experiment II (dashed functions). The effect of changing F_1 was unexpectedly large: it resulted in a dramatic increase in the variability of the voicing boundary functions for all three subjects. Only one subject, WW, showed a marked change in the level of the voicing boundary function, namely, a downward shift compared to the high F_2 condition. The other two subjects showed no clear change in the average voicing boundary, but only large discrepancies due to the variability of the low F_1 function. This is contrary to recent results of Summerfield and Haggard (1977), according to which one should have expected an increase in voicing boundaries at the lower F_1 steady-state frequency.

---

[11] Before beginning the present series of experiments, subject WW listened to the very low F_2 series of Experiment II. His results were comparable in accuracy to those of the other subjects and are shown as the dashed functions in Figure 4c.

Some statistics are shown in Table 8a. It can be seen that, in addition to the wild excursions in the voicing boundary functions, the standard deviations and standard errors were considerably larger than those of the corresponding high F_1 functions of Experiment II (see Table 5b), despite the fact that all subjects were highly practiced. Despite (or, perhaps, because of) the variability, the functions were quite reliable. Standard deviations again tended to decrease as F_2 onset increased.

Lowering F_1 not only affected the voicing boundaries but also the place boundaries. The labial-alveolar boundaries of two subjects (BHR and WW) shifted to the right, and they had unusually large standard deviations for all three subjects, even for subject WW who produced only very few velar intrusions (Table 8b). The alveolar-velar boundaries, on the other hand, were extremely sharp. (WW did not hear any velars at all in the second session, so that his alveolar-velar boundary estimate is based on the first session only.) Subjects BHR and JK produced an unusually high percentage of velar intrusions (up to 80 percent for certain F_2 onset frequencies) that extended throughout the labial category and the labial-alveolar boundary region. The (quite irregular) pattern of intrusion frequencies across stimulus series was remarkably similar for these two subjects ($r = +0.90$, $p < .001$), and so were some local features of their voicing boundary functions (see Figure 4). WW's voicing boundary function was entirely different, however. Only the steep increase across the alveolar-velar boundary was shown, as usual, by all subjects.

In summary, lowering the steady state of F_1 increased the variability and uncertainty regions of both the voicing boundary function and the labial-alveolar place boundary. Whatever caused the irregularities observed in earlier voicing boundary functions was enhanced by lowering F_1. Because of the high variability, it is difficult to decide whether the low F_1 functions merely exaggerated existing trends in the high F_1 functions, or whether they represented qualitatively different patterns.

Experiment IV: Holding F_3 Constant. As expected, the place boundaries depended on the particular fixed F_3 condition, as shown in Figure 5. Subject WW provided the cleanest data: when F_3 was rising, he divided the F_2 onset continuum fairly evenly into labials and velars. When F_3 was flat, the boundary occurred in approximately the same place, but alveolars were perceived instead of velars. When F_3 was falling, the labial-alveolar boundary shifted to the left, but labials were still heard when F_2 was rising or flat. The results of subjects BHR and JK basically agreed with this pattern. When F_3 was rising, both subjects gave a large number of alveolar responses, but, interestingly, BHR often confused alveolars with labials, while JK confused them with velars. In the other two fixed F_3 conditions, BHR and JK gave velar

Table 8: Some indices of variation and covariation (Experiment III: Low F_1).

(a) Voicing boundary functions.

Subjects	S.D. (S.E.)	$r_{I,II}$	$r_{S.D.,F_2}$
BHR	2.52 (0.47)	$+0.85^{***}$	-0.19
JK	3.61 (0.67)	$+0.75^{***}$	-0.47^{*}
WW	3.22 (0.62)	$+0.66^{***}$	-0.24

(b) Place boundary functions.

Subjects	Boundary	S.D. (S.E.)	$r_{S.D.,VOT}$
BHR	L/A	150 (18)	-0.95^{***}
	A/V	23 (5)	$+0.07$
JK	L/A	109 (16)	-0.84^{**}
	A/V	36 (7)	-0.12
WW	L/A	113 (16)	$+0.61$
	A/V[a]	31 (9)	$+0.57$

$^{*}p < .05$
$^{**}p < .01$
$^{***}p < .001$
[a] Based on one session only.

Table 9: Some indices of variation and covariation: Voicing boundary functions (Experiment IV).

Subjects	F_3 transition		
	Rising	Flat	Falling
(a) S.D. (S.E.) (msec)			
BHR	1.61 (0.33)	1.48 (0.31)	1.60 (0.33)
JK	1.72 (0.34)	1.44 (0.30)	1.50 (0.31)
WW	1.64 (0.32)	1.66 (0.33)	1.48 (0.31)
(b) $r_{I,II}$			
BHR	$+0.96^{***}$	$+0.87^{***}$	$+0.71^{***}$
JK	$+0.45^{*}$	$+0.64^{***}$	$+0.78^{***}$
WW	$+0.33$	$+0.67^{***}$	-0.18
(c) $r_{S.D.,F_2}$			
BHR	-0.08	-0.51^{**}	-0.47^{*}
JK	$+0.35$	-0.08	$+0.19$
WW	-0.49^{**}	-0.33	-0.40^{*}

$^{*}p < .05$
$^{**}p < .01$
$^{***}p < .001$

intrusions in the alveolar category. They were particularly frequent in the stimulus series with an F_2 onset frequency of 1415 Hz, and the voicing boundary functions showed corresponding peaks at this point. Both subjects heard only few labials when F_3 was falling, and then primarily at longer VOTs. (The labial-alveolar boundary of subject JK could not be determined at the shorter VOTs.)

There was little evidence that changes in the voicing boundary functions were specifically tied to place boundaries. In the rising-F_3 condition, only subject BHR showed the expected increase in voicing boundaries from labial to velar, and although a major part of this increase occurred across the place boundary, there were also systematic increases within each place category. For BHR, F_3 made little difference in the lower half of the F_2 continuum--all three voicing boundary functions were increasing. Only in the upper half of the continuum, when F_2 was steeply falling, did differences between the three functions emerge in the form of plateaus at different VOT levels. These differences seemed to be related to the percentage of velars heard, which decreased as F_3 onset increased. A somewhat similar pattern was exhibited by JK, except that his functions were flat in the lower half of the continuum and then tended to decrease. Also, he showed little difference between the rising F_3 and flat F_3 conditions, which may have been due to the mixture of alveolar and velar responses given by him in these two conditions. Subject WW showed relatively flat voicing boundary functions throughout. In the upper half of the continuum, the rising F_3 series had longer boundaries than the other two series, which is in accord with the place categories heard (velar vs. alveolar). This subject also showed some systematic differences in the lower half of the continuum, again indicating a reduction in the voicing boundary with increases in F_3 onset, but not directly corresponding to place category changes, since all these stimuli were perceived as labials.

Tables 9 and 10 again show some statistics. The standard deviations and standard errors of the voicing boundaries (Table 9a) were extremely small in all three conditions. Reliabilities were poor in some conditions, due to flatness of the functions and between-session drifts in voicing boundaries. For two subjects, standard deviations again decreased with F_2 onset frequency (Table 9c). The place boundaries, on the other hand, showed large standard deviations that increased to very large as F_3 onset frequency increased (Table 10a). This clearly demonstrates the contribution of F_3 to place distinctions: holding F_3 constant increased the uncertainty region around the place boundary. The positive relation between place boundary standard deviations and VOT, observed in earlier experiments, was not replicated here; there were even two significant negative correlations.

In summary, then, these data provide some further support for phonetic determinants of the voicing boundary, as far as the feature of velarity is concerned. Most other evidence supports the conclusions of Experiment II. The cause of the persisting irregularities in the voicing boundary functions must lie primarily in the F_2 transitions, since holding F_3 constant did not eliminate them.

Experiment V: Holding F_2 Constant. The results of this study are shown in Figure 6. One feature that immediately attracts attention is the relative

Table 10: Some indices of variation and covariation:
Place boundary functions (Experiment IV).

Subjects	F_3 transition		
	Rising	Flat	Falling
(a) S.D. (S.E.)			
BHR	90 (15)	153 (17)	179 (26)
JK	83 (13)	108 (15)	243 (26)[a]
WW	77 (13)	104 (14)	116 (18)
(b) $r_{S.D.,VOT}$			
BHR	-0.91**	-0.84**	-0.38
JK	-0.03	+0.81*	---[a]
WW	+0.23	-0.48	+0.22

*p < .05
**p < .01
***p < .001
[a] Based on only three data points.

Table 11: Some indices of variation and covariation (Experiment V).

(a) Voicing boundary functions.

Subjects	Low F_2 Onset			High F_2 Onset		
	S.D. (S.E.)	$r_{I,II}$	$r_{S.D.,F_2}$	S.D. (S.E.)	$r_{I,II}$	$r_{S.D.,F_2}$
BHR	1.56 (0.31)	+0.54*	-0.23	1.57 (0.32)	+0.86***	+0.66**
JK	1.60 (0.32)	+0.87***	+0.03	1.43 (0.30)	+0.94***	-0.06
WW	1.48 (0.31)	+0.58*	-0.17	1.33 (0.29)	+0.40	-0.26

(b) Place boundary functions.

Subjects	Low F_2 Onset		High F_2 Onset	
	S.D. (S.E.)	$r_{S.D.,VOT}$	S.D. (S.E.)	$r_{S.D.,VOT}$
BHR	214 (39)	+0.50	89 (17)	+0.75*
JK	320 (35)	+0.95***	106 (18)	+0.34
WW	248 (32)	+0.45	85 (17)	+0.78*

*p < .05
**p < .01
***p < .001

smoothness of the voicing boundary functions. This further confirms that changes in F_2, and not changes in F_3, were the source of the irregularities observed in the earlier experiments. The differences between subject BHR's voicing boundary functions in the two fixed F_2 conditions can again be rationalized in terms of the labial-velar distinction: the high F_2 function (velar category) lay above the low F_2 function (labial category), until both functions were entirely within the alveolar category, where they did not differ. A similar but less striking pattern was shown by subject WW. Subject JK, on the other hand, showed the opposite--a difference only within the alveolar category. Unless JK showed a boundary shift between the two (blocked) conditions--which would make JK's pattern similar to those of the other two subjects--JK's results suggest a direct effect of F_2 onset on the voicing boundary within the alveolar category. It should be noted that no voicing boundary function showed a major discontinuity at the place boundary. The effects observed in this experiment are thus open to an auditory interpretation: the voicing boundary tended to decrease as F_3 onset rose, and this decrease was more pronounced when the F_2 onset was high. The higher F_2 onset probably led to a higher amplitude of F_3 at onset, thus increasing its effects on the voicing boundary (and on the place boundary).

The dependency of the labial-alveolar place boundary on VOT was especially pronounced. Velar intrusions occurred only in the low F_2 condition and were relatively infrequent, more evenly distributed, and did not lead to any major peaks in the voicing boundary functions of subjects BHR and JK. Nevertheless, the distribution of these intrusions was again very similar for BHR and JK ($r = +0.77$, $p < .01$).

Table 11 again shows some statistics. The standard deviations and standard errors of the voicing boundary functions were very small and showed no clear relation to F_3 onset frequency, except for one positive correlation (Table 11a). The standard deviations and standard errors of the labial-alveolar place boundaries in the low F_2 onset condition were the largest observed in any of the present experiments (Table 11b). This shows that F_3 onset was a poor place cue when F_2 onset was low. On the other hand, the velar-alveolar boundaries in the high F_2 onset condition were much sharper, showing a much greater importance of F_3 as a place cue for this distinction. The correlations between standard deviations and VOT were again clearly positive, supporting the results of Experiments I-III.

In summary, these data provide no additional support for a phonetic dependence between voicing and place decisions. However, they show that the source of the irregularities observed in earlier voicing boundary functions was in the lower two formants.

General Discussion

The main conclusion from these experiments is that the perception of voicing in initial stops is not independent of place of articulation (in agreement with earlier studies), and, perhaps more surprisingly, that perception of place of articulation is not independent of voicing. The latter result has been independently obtained by Miller (1977), Alfonso (1977) and

Oden and Massaro (1977), and thus appears to be a reliable finding.

The dependency of the voicing boundary on place of articulation appears to be twofold: there was evidence for both phonetic and auditory effects. Consider first the auditory influences of changes in formant transitions on voicing perception. We may distinguish regular effects (such as postulated in Figure 1) from irregular effects, as primarily observed in the present experiments. It was not clear whether any regular auditory effects, that is, any truly continuous changes in the voicing boundary function, existed in the present data. The abrupt increase in the voicing boundary at the alveolar-velar place boundary could conceivably be due to a direct influence of closeness of F_2 and F_3 onsets on voicing decisions, perhaps as the two onset frequencies fall within one critical band (see Scharf, 1970). Nevertheless, a phonetic explanation of this effect seems more convincing at present. A possible regular change in the voicing boundary with F_3 onset frequency (Experiment V) likewise remains uncertain. It is fair to conclude that the present experiments have not produced clear evidence for regular auditory effects of place cues on the voicing boundary.

Irregular auditory effects, on the other hand, were ubiquitous, in the form of local peaks and troughs in the voicing boundary functions. Apparently, if the only salient voicing cue in the stimuli (VOT) was neutralized, the perceptual judgments were biased by other, marginal properties of the stimuli. As Bailey and Summerfield (1978) have suggested, any variable correlated acoustic property of the signal may become a "cue" if the major cues are neutralized. The nature of these secondary cues is of some methodological interest, but it remains to be discovered. Experiment V showed that F_3 does not play a part. Variations in the envelope of the aspiration noise are likewise ruled out as a factor by Experiment V, since the noise was just as uncontrolled there as in the other experiments. Clearly, the variations were related to changes in F_2 onset, and they were magnified when F_1 was lowered (Experiment III). F_2 may have interacted with F_1 or the harmonics of the fundamental to create minor variations in amplitude or temporal structure at stimulus onset, possibly due to limitations of the synthesizer used. A puzzle is also created by the large individual differences in the perception of these variations, as if different listeners were sensitive to different aspects of the signal. For this reason, it may prove very difficult to actually pinpoint the "cues" that led to the irregular variations in the voicing boundary functions.

The evidence for phonetic effects on the voicing boundary is threefold: the abrupt increase in the voicing boundary function at the alveolar-velar boundary (which, however, could conceivably also be auditory in nature); the peaks related to velar intrusions in two subjects; and the different values of place-conditional voicing boundaries for the same stimuli in the place boundary regions. Thus, the evidence for phonetic effects is fairly strong, especially as far as the alveolar-velar distinction is concerned. The labial-alveolar difference in voicing boundaries was less pronounced and varied from subject to subject.

The dependence of the place boundaries on VOT was the most reliable result obtained. The effect seemed regular and auditory in nature. Could VOT have had direct cue value for the place distinction? The convergence in the place boundaries with increasing VOT was shown by all subjects, although they differed widely in the shapes of their voicing boundary functions. The change in the place boundaries could be rationalized only if alveolars had shorter VOTs than labials and velars in production (which is not the case), or perhaps if all subjects exhibited shorter voicing boundaries for alveolars than for labials and velars (only subject SE showed this pattern). Therefore, the results suggest that VOT was not a direct cue for place but instead affected the perception of the transitional place cues, that is, that the effect was psychoacoustic in nature. It is important to keep in mind that the transitions of a voiced stop are not the same acoustic event as the transitions of its voiceless cognate; they differ in the source of excitation. For example, it may be that the energy in the F_3 region is relatively less salient when excited aperiodically than when excited periodically, leading to a bias against alveolars. Another psychoacoustic effect is reflected in the increase in place boundary standard deviations with VOT. The discrimination of formant transitions was somewhat more difficult when the source was aperiodic than when it was periodic, probably due the lower amplitude of the aspirated portion.

The data suggest, then, that place decisions influence voicing decisions, while voicing decisions do not influence place decisions. It need not be concluded that place decisions always precede voicing decisions, although the data would be compatible with such a fixed serial order. It seems more likely that the voicing decision is simply irrelevant to the place decision, while processing times are determined by the relative uncertainty on each dimension. The unidirectional dependency among the features at the phonetic level is in agreement with the causal relationship in articulation that leads to longer VOTs for more posterior places of articulation.

So far in this paper, phoneme recognition has been considered in terms of a simple two-stage model distinguishing (continuous) auditory and (discrete) phonetic levels (see Studdert-Kennedy, 1976). However, recently two similar models have been suggested that incorporate an intermediate stage representing the degree to which a stimulus possesses various phonetic features: the "prototype model" of Repp (1976b, 1977a) and the "fuzzy logical model" of Oden (in press; Oden and Massaro, 1977). Both models apply concepts of pattern recognition theory to speech perception and assume that the information about the characteristic auditory properties of a phoneme are stored in the form of "prototypes" in the brain. The incoming auditory information is translated by a process of "feature evaluation" (Oden, in press) into a "multicategorical code" (Repp, 1976b) that is subsequently compared to the prototypes. The prototype that matches the stimulus most closely is selected as the response. Oden's model and Repp's model differ primarily in assumptions about the nature of the matching function.

Only the fuzzy logical model has been formally tested. Oden and Massaro (1977) used bidimensional stimulus arrays similar to the present ones, but with considerably larger step sizes on both dimensions. Their model fit the data well, although it is not quite clear how surprising that result is, in view of the large number of parameters in the model. Their data replicated the present results in many respects, including the shifts in the place boundaries with VOT, and the tendency for some velar responses to occur in the region of the labial-alveolar boundary (the latter effect was not accounted for in their model). Oden's model accounts for boundary shifts in terms of properties of the perceptual prototypes for each phoneme. Thus, for example, the B-prototype is "more strongly voiced" than the G-prototype (that is, a stimulus needs a shorter VOT to be accepted as a B than to be accepted as a G), and the D-prototype is "less strongly alveolar" than the T-prototype (that is, a wider range of transition values is accepted as D-like than as T-like). This notion is very appealing, since the properties of the prototypes may be considered the listener's knowledge about the acoustic and articulatory properties of natural speech, and about the perceptual weights of VOT and formant transition cues relative to other cues (for example, bursts) that were absent in the synthetic stimuli used here and by Oden. Oden's model does not assume any direct processing interactions between the features; these dependencies are "wired in," as it were, in the prototypes.

Whether the fuzzy logical model (or its alternate, the prototype model) can account for all the present results remains to be seen. Clearly, neither model can account for some of the irregular auditory effects. There are also some discrepancies between Oden's data and the present results. For example, Oden and Massaro (1977) did not obtain any abrupt increase in the voicing boundary at the alveolar-velar boundary; changes in the voicing boundary across place categories were minimal. While the spacing of the stimuli was too coarse to follow the boundary functions in great detail, Oden's model apparently predicts a very abrupt change in the labial-alveolar place boundary with VOT, which is in disagreement with the present results. (However, it is possible that this prediction was suggested only by the schematic graphics in Figures 6 and 9 of Oden and Massaro, 1977.) At present, the model cannot explain velar intrusions, but this might be remedied by considering the place feature as two-dimensional (see Greenberg and Jenkins, 1964). It may be that the Euclidean metric of Repp's prototype model is more appropriate in this case. Data such as the present studies provide should go a long way towards further evaluation of formal models of cue integration.

REFERENCES

Abramson, A. S. and L. Lisker. (1973) Voice-timing perception in Spanish word-initial stops, *Journal of Phonetics 1*, 1-8.

Ades, T. (1976) Adapting the feature detectors for speech perception. In *New Approaches to Language Mechanisms*, ed. by E. C. T. Walker and R. J. Wales. (The Hague: North Holland).

Alfonso, P. J. (1977) Contextual influences on phonetic feature recognition in dichotic backward masking. Unpublished Ph.D. dissertation. Purdue University.

Bailey, P. J. and A. Q. Summerfield. (1978) Some observations on the perception of [s]+stop clusters. *Haskins Laboratories Status Report on Speech Research, SR-53*, vol. 1, pp.

Delattre, P. C., A. M. Liberman, and F. S. Cooper. (1955) Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America 27*, 769-773.

Draper, G. (1974) Is feature correlation directional? *Speech Perception*, Series 2, No. 3. (Belfast: Queen's University), 87-98.

Draper, G. and M. P. Haggard. (1974) Facts and artifacts in feature interdependence. *Proceedings of Stockholm Speech Communication Seminar 3*. (Uppsala: Almqvist and Wiksell), 67-75.

Eimas, P. D. and J. D. Corbit. (1973) Selective adaptation of linguistic feature detectors. *Cognitive Psychology 4*, 99-109.

Finney, D. J. (1971) *Probit Analysis*. (Cambridge: University Press).

Goldstein, L. and J. R. Lackner. (1974) Alterations of the phonetic coding of speech sounds during repetition. *Cognition 2*, 279-297.

Greenberg, J. H. and J. J. Jenkins. (1964) Studies in the psychological correlates of the sound system of American English. *Word 20*, 157-177.

Haggard, M. P. (1970) The use of voicing information. *Speech Synthesis and Perception 2*, 1-14. (Cambridge: University Press).

Harris, K. S., H. S. Hoffman, A. M. Liberman, P. C. Delattre, and F. S. Cooper. (1958) Effect of third-formant transitions on the perception of voiced stop consonants. *Journal of the Acoustical Society of America 30*, 122-126.

Hoffman, H. S. (1958) Study of some cues in the perception of the voiced stop consonants. *Journal of the Acoustical Society of America 30*, 1035-1041.

Lisker, L. (1975) Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America 57*, 1547-1551.

Lisker, L. and A. S. Abramson. (1970) The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague 1967. (Prague: Academia), 563-567.

Lisker, L., A. M. Liberman, D. Erickson, and D. Dechovitz. (1975) On pushing the voice-onset-time boundary about. *Haskins Laboratories Status Report on Speech Research SR-42/43*, 257-264.

Massaro, D. W. and M. M. Cohen. (1976) The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *Journal of the Acoustical Society of America 60*, 704-717.

Miller, J. L. (1977) The perception of voicing and place of articulation in initial stop consonants: Evidence for the nonindependence of feature processing. *Journal of Speech and Hearing Research*.

Oden, G. C. (in press) Integration of place and voicing information in the identification of synthetic stop consonants. *Journal of Phonetics*.

Oden, G. C. and D. W. Massaro. (1977) Integration of place and voicing information in identifying synthetic stop-consonant syllables. WHIPP Report 1 (Madison: University of Wisconsin).

Repp, B. H. (1975) Dichotic forward and backward 'masking' between CV syllables. *Journal of the Acoustical Society of America 57*, 483-496.

Repp, B. H. (1976a) Dichotic 'masking' of voice onset time. *Journal of the Acoustical Society of America 59*, 183-194.

Repp, B. H. (1976b) Identification of dichotic fusions. *Journal of the Acoustical Society of America 60*, 456-469.

Repp, B. H. (1976c) The voicing boundary as a function of F_2- and F_3-transitions and of fundamental frequency. *Journal of the Acoustical Society of America 60* (Suppl. no. 1), S91(A).

Repp, B. H. (1977a) Dichotic competition of speech sounds: The role of acoustic stimulus structure. *Journal of Experimental Psychology: Human Perception and Performance 3*, 37-50.

Repp, B. H. (1977b) Further observations on the function relating the voicing boundary to changes along a place continuum. *Journal of the Acoustical Society of America 61* (Suppl. no. 1), S47(A).

Repp, B. H. (1977c) 'Cross-talk' between voicing and place cues in initial stops. Journal of the Acoustical Society of America 62 (Suppl. no. 1), S79(A).

Sawusch, J. and D. B. Pisoni. (1974) On the identification of place and voicing features in synthetic stop consonants. Journal of Phonetics 2, 181-194.

Scharf, B. (1970) Critical bands. Foundations of Modern Auditory Theory vol. 1, ed. by J. V. Tobias. (New York: Academic), 157-202.

Smith, P. T. (1973) Feature-testing models and their application to perception and memory for speech. Quarterly Journal of Experimental Psychology 25, 511-534.

Stevens, K. N. and D. H. Klatt. (1974) Role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America 55, 653-659.

Studdert-Kennedy, M. (1976) Speech perception. Contemporary Issues in Experimental Phonetics. ed. by N. J. Lass. (New York: Academic).

Summerfield, A. Q. (1975a) Information-processing analyses of perceptual adjustments to source and context variables in speech. Unpublished Ph.D. dissertation. (Belfast: Queen's University).

Summerfield, A. Q. (1975b) Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables. Speech Perception, Series 2, no. 4, 61-72. (Belfast: Queen's University).

Summerfield, A. Q. and M. P. Haggard. (1977) On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. Journal of the Acoustical Society of America 62, 435-448.

II. PUBLICATIONS AND REPORTS

III. APPENDIX

PUBLICATIONS AND REPORTS

Cooper, Franklin S., Paul Mermelstein and Patrick W. Nye. (1977) Speech synthesis as a tool for the study of speech production. In *Dynamic Aspects of Speech Production*, ed. by Masayuki Sawashima and Franklin S. Cooper. (Tokyo: University of Tokyo Press), 307-322.

Dorman, Michael F., Diane Kewley-Port, Susan Brady and Michael T. Turvey. (1977) Vowel recognition: Inferences from studies of forward and backward masking. *Quarterly Journal of Experimental Psychology 29*, 483-497.

Fischer, F. William, Isabelle Y. Liberman and Donald Shankweiler. (in press) Reading reversals in developmental dyslexia. *Cortex*.

Gay, Thomas. (1977) Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America 62*, 183-193.

Gay, Thomas. (1977) Cinefluorographic and electromyographic studies of articulatory organization. In *Dynamic Aspects of Speech Production*, ed. by Masayuki Sawashima and Franklin S. Cooper. (Tokyo: University of Tokyo Press), 85-105.

Harris, Katherine S. (1977) The study of articulatory organization: Some negative progress. In *Dynamic Aspects of Speech Production*, ed. by Masayuki Sawashima and Franklin S. Cooper. (Tokyo: University of Tokyo Press), 71-82.

Katz, Leonard. (1977) Reading ability and single-letter orthographic redundancy. *Journal of Educational Psychology 69*, 653-659.

Mark, Leonard S., Donald Shankweiler, Isabelle Y. Liberman and Carol A. Fowler. (1977) Phonetic recoding and reading difficulty in beginning readers. *Memory & Cognition*, vol. 5(6), 623-629.

APPENDIX

DDC (Defense Documentation Center) and ERIC (Educational Resources Information Center) numbers SR-21/22 to SR-50:

Status Report		DDC	ERIC
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	ED-077-285
SR-33	January - March 1973	AD 762373	ED-081-263
SR-34	April - June 1973	AD 766178	ED-081-295
SR-35/36	July - December 1973	AD 774799	ED-094-444
SR-37/38	January - June 1974	AD 783548	ED-094-445
SR-39/40	July - December 1974	AD A007342	ED-102-633
SR-41	January - March 1975	AD A103325	ED-109-722
SR-42/43	April - September 1975	AD A018369	ED-117-770
SR-44	October - December 1975	AD A023059	ED-119-273
SR-45/46	January - June 1976	AD A026196	ED-123-678
SR-47	July - September 1976	AD A031789	ED-128-870
SR-48	October - December 1976	AD A036735	ED-135-028
SR-49	January - March 1977	AD A041460	ED-141-864
SR-50	April - June 1977	AD A044820	ED-144-138
SR-51/52	July - December 1977	AD A049215	ED-147-892
SR-53, v. 1	January - March 1978	**	**
SR-53, v. 2	January - March 1978	**	**

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

**DDC and/or ERIC order numbers not yet assigned.

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Haskins Laboratories, Inc.
270 Crown Street
New Haven, Connecticut 06510

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
Haskins Laboratories Status Report on Speech Research, No. 53, vol. 2, January - March, 1978

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)
Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)
Staff of Haskins Laboratories; Alvin M. Liberman, P.I.

6. REPORT DATE
March 1978

7a. TOTAL NO. OF PAGES
158

7b. NO. OF REFS
161

8a. CONTRACT OR GRANT NO.
HD-01994
V101(134)P-342
N01-HD-1-2420
RR-5596
BNS76-82023
MCS76-81034
NS13870
NS13617

9a. ORIGINATOR'S REPORT NUMBER(S)
SR-53, vol. 2 (1978)

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
None

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES
N/A

12. SPONSORING MILITARY ACTIVITY
See No. 8

13. ABSTRACT

This report (1 January - 30 April) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

-Detection Errors on The and And: Evidence for Reading Units Larger than the Word
-Some Observations on the Perception of [s] + Stop Clusters
-Perceptual Integration of Cues for Stop, Fricative and Affricate Manner
-Diachronic Tone Splits and Voicing Shifts in Thai: Some Perceptual Data
-Range Effect in the Perception of Voicing
-Note on Perceptuo-Motor Adaptation of Speech
-Interdependence of Voicing and Place Decisions for Stop Consonants in Initial Position

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

UNCLASSIFIED
Security Classification

A-31408

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Reading Errors; And and The						
Speech Perception; Fricative-stop clusters						
Speech cues; Stop, fricative, affricate, manner						
Tones; Diachronic splits and voicing in Thai						
Voicing Perception; Range effects						
Adaptation; Perceptual-Motor in speech						
Voicing and Place decisions; Interdependence						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821

UNCLASSIFIED
Security Classification

A-31409